

AD

Award Number: W81XWH-04-1-0475

TITLE: Development of a Computer-Aided Diagnosis System for Early Detection of Masses Using Retrospectively Detected Cancers on Prior Mammograms

PRINCIPAL INVESTIGATOR: Jun Wei, Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, MI 48109

REPORT DATE: June 2005

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20051018 064

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

| | | | | |
|--|--------------------------|---|---------------------------|--|
| 1. REPORT DATE (DD-MM-YYYY) 01-06-2005 | 2. REPORT TYPE Annual | 3. DATES COVERED (From - To) 1. Jun 2004 - 31 May 2005 | | |
| 4. TITLE AND SUBTITLE Development of a Computer-Aided Diagnosis System for Early Detection of Masses Using Retrospectively Detected Cancers on Prior Mammograms | | 5a. CONTRACT NUMBER | | |
| | | 5b. GRANT NUMBER W81XWH-04-1-0475 | | |
| | | 5c. PROGRAM ELEMENT NUMBER | | |
| 6. AUTHOR(S) Jun Wei, Ph.D. | | 5d. PROJECT NUMBER | | |
| | | 5e. TASK NUMBER | | |
| | | 5f. WORK UNIT NUMBER | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Michigan Ann Arbor, MI 48109 | | 8. PERFORMING ORGANIZATION REPORT NUMBER | | |
| 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012 | | 10. SPONSOR/MONITOR'S ACRONYM(S) | | |
| | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) | | |
| 12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited | | | | |
| 13. SUPPLEMENTARY NOTES | | | | |
| 14. ABSTRACT The goal of this project is to develop a computer-aided diagnosis (CAD) system for mass detection using advanced computer vision techniques that will be trained with retrospectively detected cancers on prior mammograms. The new CAD system will be combined with our existing CAD system. When fully developed, the new dual CAD system should increase the sensitivity of detecting cancers at the early stage without compromising the sensitivity for other cancers. During this project year, we have performed the following tasks: (1) begin to collect the data sets of digitized film mammograms with multiple examinations, (2) begin to develop a CAD system for mass detection on prior mammograms, and (3) perform a preliminary study to investigate the fusion scheme to combine two CAD systems. In summary, we have investigated a number of areas in CAD of mammographic masses and evaluated the new techniques for mass detection on mammograms. We have made progress in three of the tasks proposed in the project. We have found that our new computer-vision techniques can improve the performance of the CAD systems. We will continue the development of the CAD system in the coming years. | | | | |
| 15. SUBJECT TERMS Breast Cancer | | | | |
| 16. SECURITY CLASSIFICATION OF: a. REPORT U | | 17. LIMITATION OF ABSTRACT UU | 18. NUMBER OF PAGES 70 | 19a. NAME OF RESPONSIBLE PERSON 19b. TELEPHONE NUMBER (include area code) |
| b. ABSTRACT U | | | | |
| c. THIS PAGE U | | | | |

(3) Table of Contents

| | | |
|------|--|----|
| (1) | Front Cover | |
| (2) | Standard Form (SF) 298, REPORT DOCUMENTATION PAGE | |
| (3) | Table of Contents | 3 |
| (4) | Introduction | 4 |
| (5) | Body | 5 |
| | (A) Collection of a Database of Digitized Screen-film Mammograms (DFM) with Multiple Examinations | |
| | (B) Development of Single CAD System for Mass Detection on Prior DFMs. | |
| | (C) Development of a Fusion Scheme to Combine Two CAD Systems | |
| (6) | Key Research Accomplishments..... | 12 |
| (7) | Reportable Outcomes | 12 |
| (8) | Conclusions | 13 |
| (9) | References | 13 |
| (10) | Appendix | 13 |

(4) Introduction

Recent clinical studies have proved that computer-aided diagnosis (CAD) systems are helpful for improving cancer detection by radiologists on mammograms. To evaluate the effectiveness of a CAD system in detecting cancers that are likely to be missed by radiologists, one way is to study its accuracy in detecting missed cancers on prior mammograms (the mammograms in previous exams on which the cancer can be seen retrospectively). Several studies have demonstrated that CAD systems have potential ability to detect missed cancers on prior mammograms. However, the performance of a CAD system on prior mammograms is generally much lower than their performance on the current mammograms (the mammogram on which cancer is detected). Recently, one study investigated the performance change between prior mammograms and current mammograms when using the CAD system trained by current mammograms and another by prior mammograms. It was concluded that CAD schemes trained with the current mammograms do not perform optimally in detecting masses depicted on prior images and vice versa.

The goal of this proposed project is to develop a CAD system using advanced computer vision techniques to detect masses using retrospectively detected cancers on prior mammograms and incorporate the developed CAD system into our current CAD system. We hypothesize that a dual CAD system, which combines a system trained with subtle lesions retrospectively seen on prior mammograms and a system trained with cancers detected on current mammograms, should increase the sensitivity of detecting cancers at the early stage without compromising its ability to detect less subtle cancers. To accomplish this goal, we will (1) collect a large database of masses on digitized prior and current film mammograms (DFMs) for training and testing the CAD system, (2) develop single-view computer vision techniques for mass detection and classification in prior DFMs, (3) reduce false positives (FPs) by correlation of image information from two-view mammograms, (4) combine the new CAD system with our current CAD system without increase in overall FPs, and (5) perform ROC study to evaluate the effects of CAD on radiologists' accuracy in detecting subtle cancers. Although we do not plan to develop such a system for digital mammograms because there will not be enough prior digital mammograms with cancers available for the development, the general methodology developed in this study can be adapted to CAD systems for digital mammograms in the future.

At the conclusion of this project, we expect that a fully automated CAD system will be developed which can be used for detection of masses on DFMs. The general methodology developed in this study may also be adapted to develop similar software for other CAD systems. The significance of this project is that it will develop a CAD system which can further improve radiologists' accuracy in detecting breast cancers at an early stage. Since early detection and treatment can reduce breast cancer mortality rate, the CAD system will be useful for increasing the effectiveness of mammographic screening.

(5) Body

The current year (6/1/04-5/30/05) is the first year of the project. We will describe in the following details of the studies that we performed this year.

(A) Collection of a Database of Digitized Screen-film Mammograms (DFM) with Multiple Examinations

In this project year, we are collecting a data set of digitized screen-film mammogram from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. Two independent data sets of mammograms were collected for this study; one contained mammograms with masses and the other contained normal mammograms. The mass data set contained 115 cases with 115 masses. 67 of the masses are biopsy proven to be malignant and 38 to be benign. The remaining 10 masses are considered benign by long-term follow-up. Each case included the current mammograms on which the mass was detected by radiologists, and the prior mammograms obtained from previous exams. The mass set contained 230 current mammograms and 220 prior mammograms. The true location of each mass was identified by an experienced Mammography Quality Standards Act (MQSA) radiologist. The radiologist also measured the mass size and provided descriptions of the mass margin, shape, conspicuity, and breast density.

(B) Development of Single CAD System for Mass Detection on Prior DFMs.

The first study of this project is to develop a single CAD system for mass detection on prior DFMs. In this year, we have made good progress on this part. The detail is summarized in the following.

1) Data Set

Two hundred and seventy prior mammograms were collected with IRB approval from 115 patients. 22 cases with 82 mammograms have more than one year prior exams. However, 50 mammograms were negative even after radiologist's retrospective review. These 50 negative mammograms are excluded from the training and test sets so that the subtle mass set contained 220 mammograms. The true locations of the masses were identified by an experienced MQSA radiologist using the current mammogram as a reference.

2) Methods

We have previously developed a CAD system for mass detection on current DFMs. Our current CAD system consists of three main steps: detection of candidates of masses, extraction of features, and classification of mass candidates as true or false. In this study, we still used these general processing steps but new methods specifically designed for analysis of subtle masses on prior mammograms were also developed at each step. The block diagram for the detection scheme is shown in Figure 1. In the following, the new techniques developed so far are described.

For the pre-screening stage, we newly developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates a second stage in which the gray level information of the local object segmented by a region growing technique is refined by gradient field analysis. To reduce noise in the gradient calculation, the image is smoothed with a 4×4 box filter and subsampled to $400\mu\text{m} \times 400\mu\text{m}$. The gradient field analysis is applied to the smoothed image. A region of interest (ROI) of 256×256 pixels in the $100\mu\text{m} \times 100\mu\text{m}$ images is identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method in which the location of high gradient convergence is used as the starting point. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. Objects whose new gradient convergence is lower than 80% of the original value are rejected.

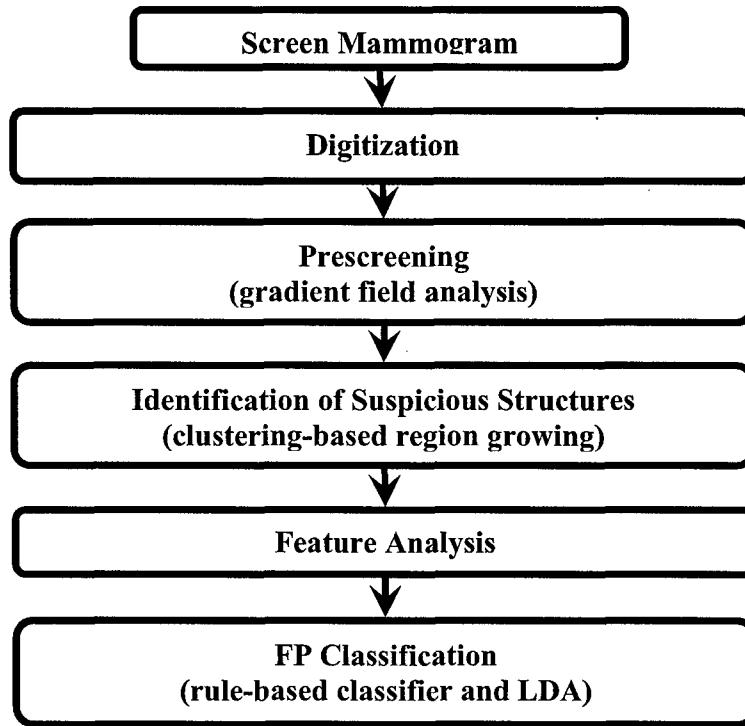


Figure 1. Schematic diagram of our single CAD system for mass detection. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

After the detection of mass candidates, the candidates will be classified as normal tissue or mass. We extracted gray level features and run length statistics analysis (RLS) texture features inside and outside of the mass region on both the original image and gradient field image. The gray level features included the contrast of the object relative to the surrounding background, the minimum and the maximum gray levels, and the characteristics derived from the gray level histogram of the object including the skewness, kurtosis, energy, and the entropy. Five RLS texture features were extracted in both the horizontal and vertical directions: short runs

emphasis, long runs emphasis, gray level nonuniformity, run length nonuniformity and run percentage. A total of 68 features were used for this CAD system.

3) Results

The detected individual objects were compared with the "truth" ROI marked by an experienced radiologist. A detected object was scored as true positive (TP) if the overlap between the bounding box of the detected object and the truth ROI was over 25%. Otherwise, it would be scored as FP. The 25% threshold was selected as described in our previous study. We randomly separated the data set into two independent subsets: one contained 57 cases with 109 prior images and the other contained 58 cases with 111 prior images. Cross validation was used for training and testing the algorithms. The training included selecting proper parameters for single CAD systems. Once the training with one subset was completed, the parameters were fixed for testing with the other subset. The training and test subsets were switched and the training process was repeated. The overall detection performance was evaluated by combining the performances for the two test subsets. The detection performance of the CAD system was assessed by free response ROC (FROC) analysis. The average FROC curves were obtained by averaging the FP/images at the corresponding sensitivities from the two mass test subsets. Figure 2 shows the average FROC curves of our CAD system for detection of subtle masses on prior mammograms by using image-based scoring and case-based scoring.

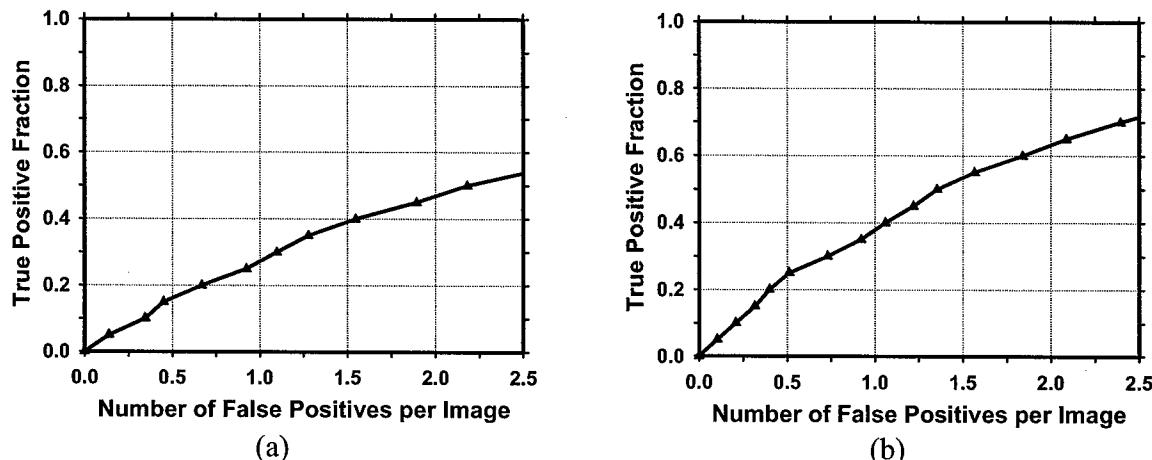


Figure 2. The average FROC curves of our CAD system for detection of masses on prior mammograms. (a) image-based FROC, (b) case-based scoring.

We have recently applied these improved techniques to mass detection on full field digital mammograms (FFDM) and found that they were useful for improving the accuracy of mass detection on current mammograms as well. A journal article with the results on FFDMs is now in press in the Medical Physics Journal (enclosed as Appendix).

(C) Development of a Fusion Scheme to Combine Two CAD Systems

The second study performed in this project year is to develop a fusion scheme to combine a CAD system optimized with "average" masses with another CAD system optimized with

“subtle” masses. In this project year, we performed a preliminary study to investigate an artificial neural network (ANN) for information fusion. Our preliminary results were presented at the RSNA meeting in 2004 and the SPIE meeting in 2005. The study is summarized in the following.

1) Data Set

Two independent mammogram data sets which included a data set with masses and a data set with normal mammograms were used in this study. All mammograms were collected from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. The abnormal data set contained 115 cases with 115 masses, in which 105 masses are biopsy-proven and 10 masses are with long-term follow-up. Each case included the current mammograms and the prior mammograms. The masses on the current mammograms are referred to as the average mass set. The masses on the prior mammograms represent a subtle mass set. The normal data set contained 260 mammograms of 65 patients. Each case contains CC view and MLO view of both breasts of the patient. The abnormal data set was used to estimate the detection sensitivity and the normal data set was used for estimating the FP rate. Figure 3(a) and (b) shows the distributions of mass sizes in the abnormal data set. To train the dual system, we randomly split the abnormal data set into two independent subsets, each of which contained about half of the current mammograms and half of the priors in the abnormal data set. One subset was used training and the other was used for testing and vice versa.

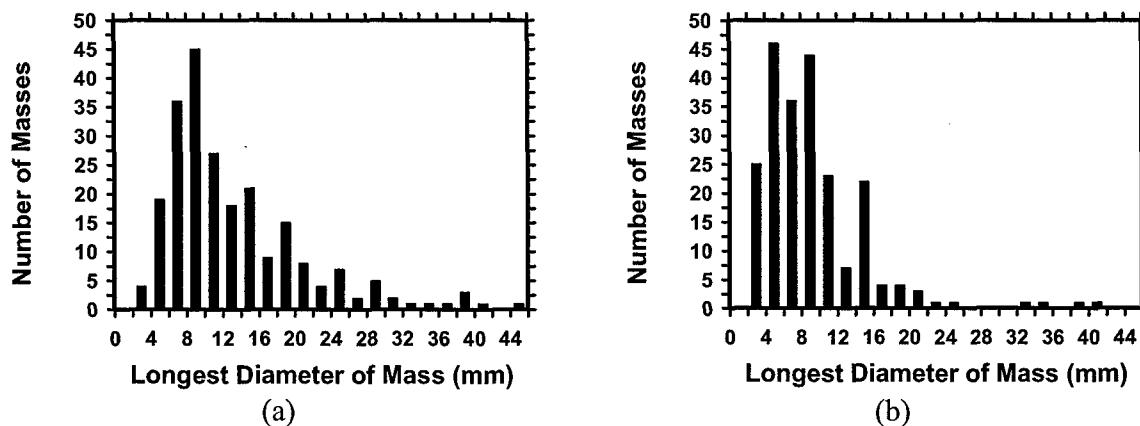


Figure 3. The distribution of mass sizes in our data set. (a) The “average” mass set, (b) The “subtle” mass set.

2) Methods

Two CAD systems for mass detection were trained; one system was trained with the current mammograms and the other trained with the prior mammograms. We have investigated methods to merge the information from the two CAD systems. In this study, a feed-forward backpropagation artificial neural network (BP-ANN) was trained to classify the masses from normal tissues by combining the output information from the two single CAD systems. In this ANN, the nodes are organized in an input layer, an output layer, and one hidden layers. The two

LDA discriminant scores from the two CAD systems were used as input to the BP-ANN. The BP-ANN has two input nodes, a single hidden layer with 3 hidden nodes, and one output node. The nodes are interconnected by weights and information propagates from one layer to the next through a log-sigmoidal transform function. The learning of the ANN is a supervised process in which known training cases are input to the ANN. The performance function for the feedforward network was the mean-square error which was the average squared error between the network outputs and the target values over all training samples. The gradient of the performance function was used to determine how to adjust the weights to minimize the error. The gradient is determined using an iterative backpropagation procedure which involves performing computations backward through the network.

To train the BP-ANN, we used a 3-fold cross-validation method within the training subset. We randomly separated the entire training subset to three independent groups. The mammograms which belong to the same case were grouped into the same subset. Three individual training processes were performed to determine the best training parameters. During each individual training process, two of the three groups were used to train the BP-ANN and the left-out group was used to test the performance of ANN. The Az value of test results obtained from ROC analysis was used as the performance index for each training process. The trained ANN with the largest Az value was considered to be the best and chosen to merge the information from the two single systems.

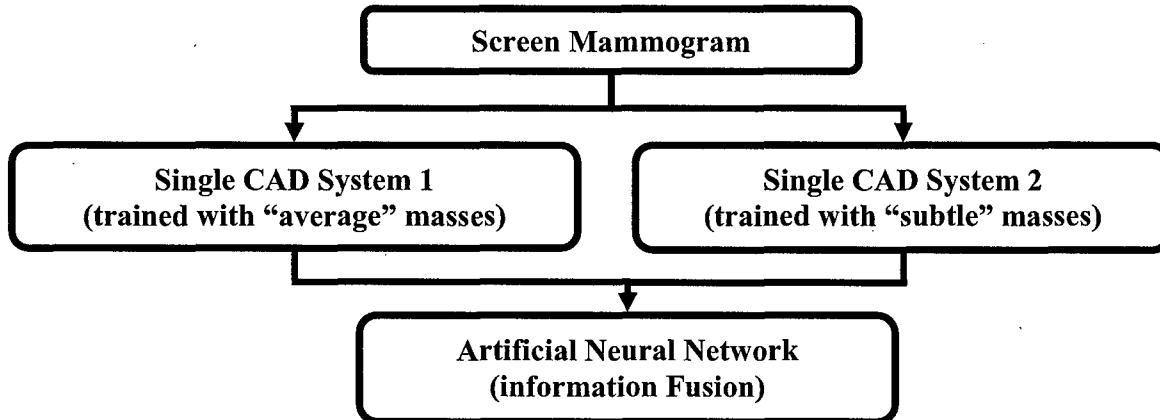


Figure 4. Schematic diagram of proposed dual CAD system for mass detection. A BP-ANN is used for information fusion. During the training procedure, both the “average” masses on the current mammograms and the “subtle” masses on the prior mammograms are used to train the ANN classifier. During the testing of our dual system, the unknown test mammograms will be the inputs for both trained single CAD system. No prior mammogram is needed during testing.

To test the dual system, the two trained CAD systems, one trained with the average mass set and the other with the subtle mass set, were applied in parallel to each single “unknown” mammogram in the independent test subset. No prior mammogram is needed during testing. The

block diagram for the dual system is shown in Figure 4.

3) Results

Since we used two-fold cross validation to train the dual CAD system, we obtained two sets of trained parameters. We applied the two dual CAD systems separately to the normal data set for FP detection. The number of FP marks produced by the algorithm was determined by counting the detected objects on these normal cases only. The mass detection sensitivity was determined by counting only the masses on each of the test mass subsets. The combination of the sensitivity from each of the test mass subsets and the FP rate from the normal data set at the corresponding detection thresholds resulted in a test FROC curve. The two test FROC curves were then averaged, as described above, to obtain an overall FROC curve quantifying the test performance of the CAD system. Figure 5 shows the comparison of the average FROC curves with the FP rates estimated from the normal data set.

In this study, we have 67 malignant cases. Figure 6 shows the average FROC curves for detection of malignant masses only by using the single CAD system and the dual system. The single CAD system trained with average masses was used and the FP rate was estimated from the mammograms without masses. In this case, the dual CAD system achieved a case-based sensitivity of 80%, 85%, and 90% at 0.6, 0.9, and 1.2 FP marks/image, respectively, compared with 1.2, 1.6, and 2.0 FP marks/image on the single CAD system.

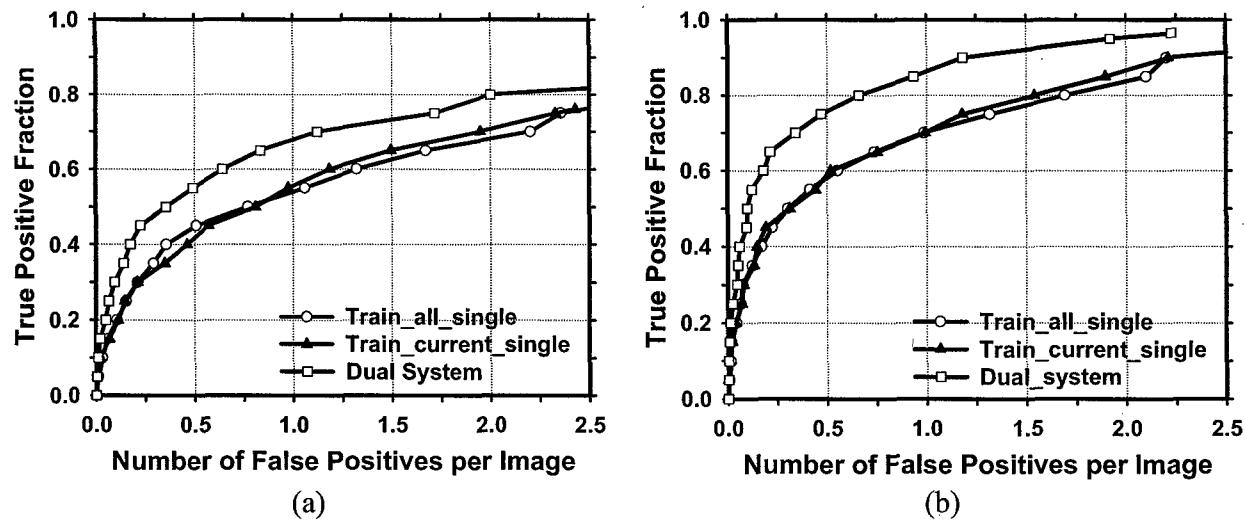


Figure 5. Comparison of the average test FROC curves obtained by averaging the FROC curves from the two independent mass subsets containing average masses. Three CAD systems were compared: a single CAD system trained with “average” masses only, a single CAD system trained with “average” and the “subtle” masses, and the dual CAD system. The FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

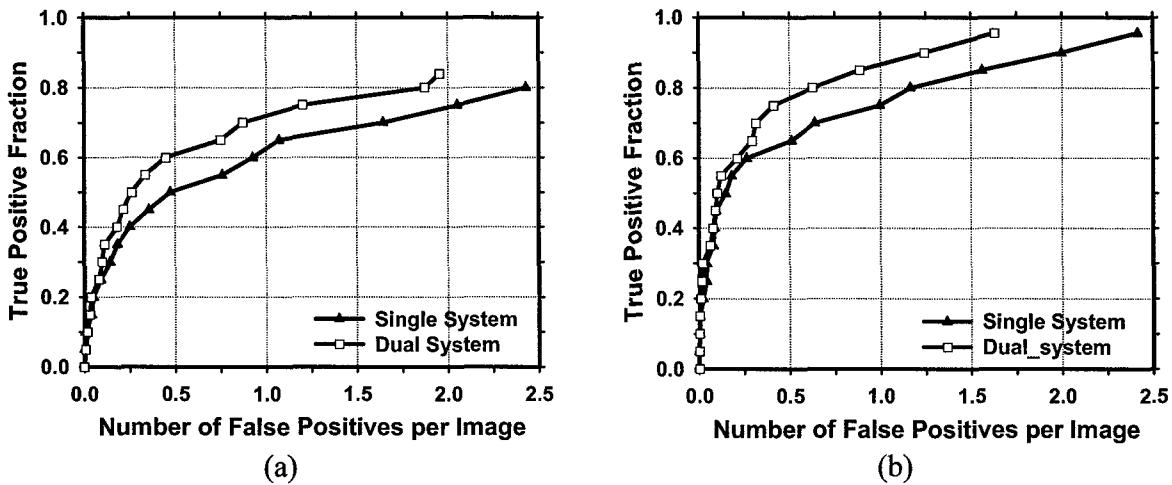


Figure 6. Comparison of the average test FROC curves for the malignant masses only in the average data set by using the single CAD system and the dual CAD system. The single system was trained with the average masses only and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

4) Discussion

For the comparison of the different systems, we analyzed the false negatives (FN) on the single CAD system and the dual CAD system when the test subsets with “average” masses on current mammograms were used. It was found that the FN rates of the single CAD system trained with “average” masses, the single CAD system trained with “subtle” masses and the dual system are 23.9% (55/230), 28.3% (65/230) and 16.5% (38/230), respectively. The number of corresponding FNs for the two single systems is 29. By using the dual system, 53 masses which were FNs for either single system can be detected. However, the masses which cannot be detected by both of the single CAD systems were also missed by the dual CAD system.

Table 1. Estimation of the statistical significance in the difference between the FROC performance of the single CAD system and the dual system. The FROC curves with the FP rates obtained from the normal data set were compared.

| | A_1 (AFROC) | | FOM (JAFROC) | |
|---------------|---------------|---------------|---------------|---------------|
| | Test subset 1 | Test subset 2 | Test subset 1 | Test subset 2 |
| Single system | 0.45 | 0.44 | 0.48 | 0.48 |
| Dual system | 0.55 | 0.53 | 0.60 | 0.56 |
| P values | 0.0004 | 0.0156 | 0.00002 | 0.007 |

In order to compare the FROC curves from the single CAD system and the dual CAD system, we used a JAFROC method and an AFROC method. Table I summarized the results of statistical significance testing using the methods. The results indicate that the ANN fusion scheme can provide significant improvement in the accuracy of the mass detection CAD system in comparison with that of a single CAD system. Our proposed dual CAD system is thus a promising approach to improving mass detection on mammograms. We will continue to improve the various stages of the dual CAD system in the coming year.

(6) Key Research Accomplishments

- Begin to collect the data sets of digitized film mammograms with multiple examinations. (Task 1).
- Develop a CAD system for mass detection on prior mammograms. (Task 2).
- Investigate the use of BPANN as a fusion scheme to combine two CAD systems. (Task 4).

(7) Reportable Outcomes

As a result of the support by the USAMRMC BCRP grant, we have conducted studies to develop a computer-aided diagnosis system for early detection of masses using retrospectively detected cancers on prior mammograms. We have presented the results of these investigations in this project year and a journal article was accepted for publication by the Medical Physics Journal.

Journal Articles:

1. Wei J, Sahiner B, Hadjiiski, Chan HP, Petrick N, Helvie MA, Roubidoux MA, Ge J, Zhou C. Computer-aided detection of breast masses on full field digital mammograms. Medical Physics (in press).

Conference Proceeding:

1. J. Wei, B. Sahiner, L. M. Hadjiiski, HP. Chan, M. A. Helvie, M. A. Roubidoux , N. Petrick, C. Zhou, J. Ge, "Computer aided detection of breast masses on mammograms: performance improvement using a dual system", Proc. SPIE Int. Soc. Opt. Eng., Vol. 5747, pp. 9-15, 2005.

Conference Presentation:

1. J. Wei, B. Sahiner, L. M. Hadjiiski, HP. Chan, M. A. Helvie, M. A. Roubidoux, "A dual computer-aided detection (CAD) system for improvement of mass detection on mammograms", Presentation at the 90th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, November 28-December 3, 2004.

2. J. Wei, B. Sahiner, L. M. Hadjiiski, HP. Chan, M. A. Helvie, M. A. Roubidoux , N. Petrick, C. Zhou, J. Ge, "Computer aided detection of breast masses on mammograms: performance improvement using a dual system", Presentation at SPIE Medical Imaging 2005, San Diego, CA, February 14-18, 2005.
3. J. Wei, B. Sahiner, L. M. Hadjiiski, HP Chan, M. A. Helvie, M. A. Roubidoux, "A dual system for improvement of computer-aided mass detection on mammograms", Poster presentation at the Era of Hope Meeting, U. S. Army Medical Research and Materiel Command, Department of Defense, Breast Cancer Research Program, Phil., June, 2005.

(8) Conclusions

During this project year, we have developed a CAD system for mass detection on prior mammograms and also developed a fusion scheme to combine two CAD systems. When the single CAD system trained on the average mass data set was applied to the test subset, the FP rates were 2.2, 1.9 and 1.5 per image at the case-based sensitivities of 90%, 85% and 80%, respectively. With the dual CAD system, the FP rates were improved to 1.2, 0.9 and 0.7 per image at the same case-based sensitivities. In order to compare the FROC curves from the single CAD system and the dual CAD system, we used a JAFROC method. It was found that the differences between the dual CAD system and the single CAD system on the two test subsets were statistically significant ($p<0.05$). Our results indicate that the dual CAD system could achieve a higher accuracy than the single CAD system.

(9) References

NA

(10) Appendix

Copies of the following publications are enclosed with this report:

Journal Articles:

1. Wei J, Sahiner B, Hadjiiski, Chan HP, Petrick N, Helvie MA, Roubidoux MA, Ge J, Zhou C. Computer-aided detection of breast masses on full field digital mammograms. Medical Physics (in press).

Conference Proceeding:

1. J. Wei, B. Sahiner, L. M. Hadjiiski, HP. Chan, M. A. Helvie, M. A. Roubidoux , N. Petrick, C. Zhou, J. Ge, "Computer aided detection of breast masses on mammograms: performance improvement using a dual system", Proc. SPIE Int. Soc. Opt. Eng., Vol. 5747, pp. 9-15, 2005.

Computer Aided Detection of Breast Masses on Full Field Digital Mammograms

Jun Wei, PhD

Berkman Sahiner, PhD

Lubomir M. Hadjiiski, PhD

Heang-Ping Chan, PhD

Nicholas Petrick, PhD

Mark A. Helvie, MD

Marilyn A. Roubidoux, MD

Jun Ge, PhD

Chuan Zhou, PhD

Department of Radiology

University of Michigan, Ann Arbor

Correspondence :

Jun Wei, Ph.D.

Department of Radiology

University of Michigan

CGC B2103

1500 E. Medical Center Drive

Ann Arbor, MI 48109

Phone: 734-647-8553

Fax: 734-615-5513

Email: jvwei@umich.edu

Running head: Computer-aided detection of masses on digital mammograms

Abstract

We are developing a computer-aided detection (CAD) system for breast masses on full field digital mammographic (FFDM) images. To develop a CAD system that is independent of the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM image as input and developed a multi-resolution preprocessing scheme for image enhancement. A two-stage prescreening method that combines gradient field analysis with gray level information was developed to identify mass candidates on the processed images. The suspicious structure in each identified region was extracted by clustering-based region growing. Morphological and spatial gray-level dependence (SGLD) texture features were extracted for each suspicious object. Stepwise linear discriminant analysis (LDA) with simplex optimization was used to select the most useful features. Finally, rule-based and LDA classifiers were designed to differentiate masses from normal tissues. Two data sets were collected: a mass data set containing 110 cases of two-view mammograms with a total of 220 images, and a no-mass data set containing 90 cases of two-view mammograms with a total of 180 images. All cases were acquired with a GE Senograph 2000D FFDM system. The true locations of the masses were identified by an experienced radiologist. Free-response receiver operating characteristic (FROC) analysis was used to evaluate the performance of the CAD system. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% at 0.72, 1.08, and 1.82 false positive

(FP) marks/image on the mass data set. The FP rates on the no-mass data set were 0.85, 1.31, and 2.14 FP marks/image, respectively, at the corresponding sensitivities. This study demonstrated the usefulness of our CAD techniques for automated detection of masses on FFDM images.

Keywords: Computer-Aided Detection, Full Field Digital Mammogram (FFDM), Multi-resolution Image Enhancement, Gradient Field Analysis, Stepwise Linear Discriminant Analysis

I. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 to 55 years of age¹. It has been reported that early diagnosis and treatment can significantly improve the chance of survival for patients with breast cancer²⁻⁴. Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially⁵⁻⁸. Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the sensitivity of mammography^{9,10}. Computer-aided lesion detection can be used during screening to reduce oversight of suspicious lesions that warrant further work-up. Computer-aided lesion characterization can assist in the estimation of the likelihood of malignancy of lesions by using image and/or other information during the diagnostic stage. The majority of studies to-date shows that CAD can improve radiologists' lesion detection sensitivity¹¹⁻¹⁶, although Gur et al.¹⁷ found that CAD had no significant effect on the radiologists in their academic setting when they averaged the results from both low-volume and high-volume radiologists. Further analysis of Gur's data by Feig et al.¹⁸ indicated that the 17 low-volume radiologists in Gur's study achieved similar increase in sensitivity as reported in other studies. The outcome of CAD studies therefore depends on the study design and data analysis.

A number of investigators have reported CAD algorithms for detection of masses on mammograms. Their approaches to prescreening of mass candidates were based primarily on mass characteristics including: (1) asymmetric density between left and right mammograms¹⁹⁻²², (2) texture^{23,24}, (3) spiculation^{25,26}, (4) gray level contrast²⁷⁻³¹, and (5) gradient³². Some of these approaches were refined with a combination of the mass characteristics. Feature classifiers were then used to further differentiate masses from normal breast tissues.

Most mammographic CAD algorithms developed so far are based on digitized screen-film mammograms (SFMs). In the last few years, FFDM technology has advanced rapidly because of the potential of digital imaging to improve breast cancer detection. Several manufacturers have obtained clearance from FDA for clinical use to date. It is expected that FFDM detectors will provide higher signal-to-noise ratio (SNR) and detective quantum efficiency (DQE), wider dynamic range, and higher contrast sensitivity than digitized mammograms. The spatial resolution of digital detectors may also be different from that of digitized SFMs even when their pixel pitches are equal. Zheng et al. investigated the performance of their CAD system on mass detection that was developed for SFMs and modified for FFDMs³³. Their preliminary results on a small data set showed that it achieved 60% sensitivity at 2.47 false positives (FPs)/image. It is expected that proper adaptation based on the imaging characteristics of FFDMs and re-training of the CAD system with FFDMs would

improve the performance. Because of the higher SNR and linear response of digital detectors, there is also a strong potential that more effective feature extraction techniques can be designed to optimally extract signals from the image and improve the accuracy of CAD. Several commercial CAD systems already obtained FDA approval for use with FFDMs. The commercial CAD systems generally reported similar performance on FFDMs and SFMs. However, their study was not reported in peer-reviewed journals so that the data set and algorithm are unknown. Recently, an assessment study³⁴ to compare the performance of two commercial and one research CAD systems for SFMs showed that their mass detection sensitivities ranged from 67% to 72% and the FP rates ranged from 1.08 to 1.68 per four-view examinations. The differences in sensitivities were not significant whereas the differences in the FP rates were significant, depending on the examinations and CAD systems used³⁴.

We have developed a CAD system for the detection of masses on SFMs in our previous studies^{30,35,36}. We are developing a mass detection system for mammograms acquired directly by an FFDM system. In this study, we adapted our mass detection system developed for SFMs to FFDMs by optimizing each stage and retraining. In an effort to develop a CAD system that is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM as input and developed a multi-resolution preprocessing scheme for image enhancement. A new technique was also designed for prescreening of mass candidates on the preprocessed

images.

II. MATERIALS AND METHOD

2.1 Data Sets

The mammograms were collected from patient files in the Department of Radiology with Institutional Review Board (IRB) approval. Digital mammograms at the University of Michigan are acquired with a GE Senographe 2000D FFDM system. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of $100\mu m \times 100\mu m$ and 14 bits per pixel. In this study, we used two data sets: a mass set containing FFDMs with malignant or benign masses and a no-mass set containing FFDMs without masses. The no-mass set was obtained from microcalcification cases collected for the development of our microcalcification CAD systems. The cases were included as normal, with respect to masses, only if they were verified to be free of masses by an experienced Mammography Quality Standards Act (MQSA) radiologist. Our mass detection system aims at application to screening mammography so that the mass cases, regardless of malignant or benign, are considered positive. All cases had two mammographic views, the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral (LM or ML) view. The mass set contained 110 cases with a total of 220 images. The no-mass set contained 90 cases with a total of 180 images. The mass

data set was used to estimate the detection sensitivity and the no-mass data set was used for estimating the FP rate. There were a total of 110 biopsy-proven masses in the mass data set. Eighty seven of the masses were benign and 23 of the masses were malignant. An MQSA radiologist identified the locations of the masses, measured the mass sizes as the longest dimension seen on the two-view mammograms, provided descriptors of the mass shapes and mass margins, and also provided an estimate of the breast density in terms of BI-RADS category. Figure 1 shows the information of our data set which includes the distributions of mass sizes, mass shapes, mass margins and breast density.

2.2 Methods

Our CAD system consists of five processing steps: (1) preprocessing by using multi-scale enhancement, (2) pre-screening of mass candidates, (3) identification of suspicious objects, (4) feature extraction and analysis, and (5) FP reduction by classification of normal tissue structures and masses. The block diagram for the detection scheme is shown in Figure 2. These steps are described in more detail below.

We randomly separated the mass data set into two independent, equal sized subsets. Each subset contained 55 cases with 110 images. Cross validation was used for training and testing the algorithms. The training included selecting the preprocessing Laplacian pyramid reconstruction weights, adjusting the filter weights for prescreening and clustering, determining

thresholds for rule-based classification, and selecting morphological and texture features and classifier weights. Once the training with one subset was completed, the parameters and all thresholds were fixed for testing with the other subset. The training and test subsets were switched and the training process was repeated. The overall detection performance was evaluated by combining the performances for the two test subsets. The trained algorithms with the fixed parameters were also applied to the no-mass mammograms to estimate the FP rate in screening mammograms.

A. Preprocessing

FFDMs are generally preprocessed with proprietary methods by the manufacturer of the FFDM system before being displayed to readers. The image preprocessing method used depends on the manufacturer of the FFDM system. To develop a CAD system that is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we use the raw FFDM as input to our CAD system. We developed a multi-scale preprocessing scheme for image enhancement.

Multi-scale methods have been used for contrast enhancement of medical images. Since a multi-scale method uses the information from a large number of frequency channels extracted from the image adaptively, it is more flexible and versatile than the commonly used enhancement methods, such as unsharp masking, which uses a small number of frequency channels. Two

types of multi-scale methods have been used as the preprocessing methods for the contrast enhancement of mammograms: the wavelet method and the Laplacian pyramid method³⁷. A previous study has shown that, for the purpose of image enhancement, using a Laplacian pyramid method is advantageous compared to using the fast wavelet transformation which introduces visible artifacts³⁸. In this project, therefore, we chose the Laplacian pyramid method as our preprocessing method.

A flowchart of our preprocessing method is shown in Figure. 3. In brief, the mammogram is first segmented automatically into the background and the breast region. Second, a logarithmic transform is applied to the breast image. The Laplacian pyramid method is used to decompose the breast image into multi-scales. A nonlinear weight function based on the pixel gray level from each of the low-pass components is designed to enhance the high-pass components.

Since the contrast between the breast and the background in a raw FFDM is high, a two-step algorithm was developed for the segmentation of breast region. First, Otsu's method³⁹ is used to calculate a threshold and binarize the original image. Second, an 8-connectivity labeling method is used to identify the connected regions below the threshold on the binary image. The region with the largest area will be considered to be the breast region.

Clinical mammograms are usually viewed in a negative mode of the raw images. In

order to process an image with the same format as the clinical mammograms, we first use an inverted logarithmic function⁴⁰ to transform the raw data. A multi-resolution method is then used to enhance the log-transformed image. The inverted logarithmic function for signal transfer can be expressed as

$$S_x = \ln(\frac{X_{\max}}{X}) \quad (1)$$

where X is the gray level of the raw data, X_{\max} is the maximum of the 14-bit digital gray scale number (i.e., 16383). The transformed image is then linearly scaled to 12-bit pixel values.

The Laplacian pyramid decomposition is a multi-scale method that was first introduced as an image compression technique³⁷. We previously evaluated the effect of Laplacian pyramid data compression on the detection of microcalcifications on digitized mammograms⁴¹. An illustration of a Laplacian decomposition tree is shown on the left side of Figure 4. The Laplacian pyramid is a sequence of error images L_0, L_1, \dots, L_n . Each is the difference between two consecutive levels of the Gaussian pyramid G_0, G_1, \dots, G_n , where G_0 is the original image. Each subsequent level of the Gaussian pyramid in the decomposition tree is generated by convolution of the image at the previous level with a 5×5 kernel, $w(m, n)$, that has weights of 0.4 at the center, 0.25 at the eight nearest neighbors of the center, and 0.05 at the 16 peripheral pixels, and then down-sampled by a factor of 2, as described in Eq. (4). The decomposition of the image from level k to level $k+1$ can be expressed mathematically by the following equation:

$$L_k = G_k - \text{Expand}(G_{k+1}) \quad (2)$$

where

$$\text{Expand}(G_{k+1}) = 4 \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) \cdot G_{k+1}\left(\frac{i-m}{2}, \frac{j-n}{2}\right) \quad (3)$$

$$G_k(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) G_{k-1}(2i+m, 2j+n) \quad (4)$$

The original image can be recovered by following the Gaussian reconstruction tree shown on the right side of Figure 4 if no enhancement is applied to the Laplacian pyramid. At a given level of the Gaussian reconstruction tree, the image is expanded (convolved and upsampled), as shown in Eq. (3), and then added to the Laplacian error image of the corresponding level. Details of the decomposition and reconstruction processes can be found in the literature³⁷.

We enhance the reconstructed image to facilitate mass detection. The image at each level of the Laplacian pyramid that corresponds to a bandpass image is mapped by a nonlinear function. In this study, we use a nonlinear function that incorporates the information from each bandpass image. A Gaussian pyramid expansion is then used to reconstruct the image from the low pass components and the enhanced bandpass components, as shown in Figure 4. The reconstruction scheme is defined by

$$r(k) = \alpha \cdot \text{Expand}(G_{k+1}) + \beta \cdot (\text{Expand}(G_{k+1}))^p \cdot L_k \quad (5)$$

where α , β , and p are constant values in the range of 0.2 to 2.0 experimentally chosen for each frequency level.

Figures 5(a) and 5(b) show an example of a GE raw image and its processed image provided by the GE FFDM system. The histograms of the raw image and the processed image are shown next to the corresponding images. An example of the processed image using our multi-resolution enhancement method and the corresponding histogram are shown in Figure 5(c).

B. Prescreening and segmentation of suspicious objects

In our previous CAD system developed for digitized SFMs, an adaptive density-weighted contrast enhancement (DWCE) filter³⁵ was developed for prescreening. Although the DWCE filter using the gray level information can identify the suspicious locations of masses on mammograms with high sensitivity, the prescreening objects often include a large number of enhanced normal breast structures.

In this study, we investigated the use of a new method that combines gradient field information and gray level information to detect mass candidates on FFDMs. Gradient field information is commonly used in computer vision or other fields to extract objects or intensity field distributions. Kobatake et al.⁴² designed a filter, referred to as an iris filter, to calculate the convergence of gradient index around each pixel on SFMs which provided shape information for detection of masses. An extension of the iris filter, referred to as an adaptive ring filter, was developed by Wei et al.⁴³ for detection of lung nodules on chest x-ray images. In this study, we have developed a two-stage gradient field analysis method which uses not only the shape

information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis.

To reduce noise in the gradient calculation, the image is smoothed with a 4×4 box filter and subsampled to $400\mu m \times 400\mu m$. The gradient field analysis is applied to the smoothed image. At each pixel $c(i)$ within the breast, concentric annular regions centered at $c(i)$ with an average radius, $R(k)$, of k pixels from $c(i)$ and a radial width of 4 pixels are defined within a circular region of about 12 mm in radius. The gradient vector at each pixel $p(j)$ within an annular region is computed and the gradient direction is obtained by projecting the gradient vector to the radial direction vector from $c(i)$ to $p(j)$. The average gradient direction over an annular region at the average radius $R(k)$ is calculated as the mean of the gradient directions over pixels on three adjacent annular regions $R(k-1)$, $R(k)$, and $R(k+1)$. Finally, the gradient field convergence at $c(i)$ was determined as the maximum of the average gradient directions among all annular regions. A region of interest (ROI) of 256×256 pixels in the $100\mu m \times 100\mu m$ images is identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method⁴⁴ in which the location of high gradient convergence is used as the starting point. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient

convergence at the center location of the ROI is recalculated within the segmented object.

Objects whose new gradient convergence is lower than 80% of the original value are rejected.

After prescreening, the suspicious objects are identified by using a two-stage segmentation method. First, the background-corrected ROI was weighted by a Gaussian function with $\sigma=256$ pixels. Then, a k-means clustering using the pixel values in a background-corrected image and a Sobel filtered image as features is used to find the object. Figures 6(a) and 6(b) show the initial detection locations and the grown objects, respectively, obtained by prescreening the mammogram shown in Figure 5(c).

C. Feature extraction and FP reduction

FP classification in our mass detection system is accomplished by a three-stage classification scheme^{36,44}. For each suspicious object, eleven morphological features are extracted. Rule-based classification and an LDA classifier using all 11 morphological features as input predictor variables are trained to remove the detected structures that are substantially different from breast masses. The training data set alone was used for training the classification rules and the weights of the LDA classifier. After morphological classification, global and local multi-resolution texture analysis⁴⁵ are performed in each remaining ROI by using the spatial gray level dependence (SGLD) matrix. Briefly, the wavelet transform is employed to decompose an ROI into three levels for global texture analysis. Thirteen types of texture features^{44,46} are

extracted from each ROI. Each feature is calculated at 14 pixel distances and 2 angular directions. A total of 364 features (13 texture measures \times 14 distances \times 2 directions) are extracted from global texture analysis. Local texture features are extracted from the local region containing the detected object (object region) and the peripheral regions within each ROI. A total of 208 features (104 features from the object region and 104 features from the peripheral regions) are extracted. The third-stage FP reduction using the texture features is described next.

D. Texture classification of masses and normal tissue

In order to obtain the best texture feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection with stepwise linear discriminant analysis (LDA) was applied. At each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. The optimization procedure used a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for excluding features that had high correlation with the features already in the selected pool. Since the appropriate values of F_{in} , F_{out} and T were unknown, we examined a range of F_{in} , F_{out} , and T values using an automated simplex optimization method. For a given combination of F_{in} , F_{out} , and T values, the algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To evaluate the classifier performance, the test discriminant

scores from the left-out cases were analyzed using receiver operating characteristic (ROC) methodology⁴⁷. The discriminant scores of the mass and normal tissue were used as the decision variable in the LABROC program, which fits a binormal ROC curve based on maximum likelihood estimation. The accuracy for classification of mass and normal tissue was evaluated as the area under the ROC curve, A_z . The test A_z for the left-out cases in the leave-one-out resampling within the training subset was used as a figure of merit to guide the simplex algorithm to search for the best set of F_{in} , F_{out} , and T values within the parameter space. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased⁴⁸. However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one training case different from the other cycles. In order to obtain a single trained classifier to apply to the test subset, a final stepwise feature selection was performed with the entire training subset and a set of F_{in} , F_{out} , and T thresholds chosen from the output of simplex training process. This set of F_{in} , F_{out} , and T thresholds was chosen based not only on the test A_z values, which were generated when the simplex procedure was searching through the parameter space, but also on the average number of features selected. The appropriate thresholds were chosen as a balance between keeping the number of selected features small and a relatively high classification accuracy by LDA. The chosen thresholds were then applied to the entire training subset to obtain the final

set of features using stepwise feature selection and estimate the weights of the LDA. The LDA classifier with the selected feature set were then fixed and applied to the test subset. The test subset was independent of the training subset as described in Section II.2.2 and was not used in the leave-one-case-out classifier training process describe above.

E. Evaluation methods

The detected individual objects were compared with the “truth” ROI marked by an experienced radiologist. A detected object was scored as true positive (TP) if the overlap between the bounding box of the detected object and the truth ROI was over 25%. Otherwise, it would be scored as FP. The 25% threshold was selected as described in our previous study³⁶. The detection performance of the CAD system was assessed by free response ROC (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 220 masses. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection; the sensitivity was thus calculated relative to 110 masses.

Figure 6(c) shows an example of the final detected objects and Figure 6(d) shows the locations of these objects superimposed on the mammogram.

To evaluate the effect of the preprocessing methods on mass detection, we also trained a CAD system using the GE processed image as input. This CAD system used the same methods as those described above for the raw images except that the Laplacian pyramid preprocessing step was not applied to the GE processed image, and that the prescreening and feature classifiers were retrained specifically for the GE processed images to obtain the best performance. The training and test subsets contained the same corresponding cases as for the raw image subsets. The training and testing were performed using the cross validation method as described above. The performance of the CAD system using the GE processed images was quantified by the average test FROC curve and compared with that using the raw images.

III. RESULTS

With raw images as input and Laplacian pyramid enhancement, our CAD system using the two-stage gradient field analysis detected 92.7% (204/220) of the masses with an average of 18.9 (4152/220) objects/image at the prescreening stage, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone. After FP reduction using the rule-based and linear classifier based on morphological features, there were a total of 3412 mass candidates (15.5 objects/image) at a sensitivity of 90.5% (199/220).

The texture-based LDA classifier for FP reduction was designed with stepwise feature selection and simplex optimization. The most effective subset of features from the available

feature pool was selected for each of the training subsets during the training procedure. Twenty (11 global and 9 local) and 19 (12 global and 7 local) texture features were selected from the two independent training subsets, respectively. The test ROC curves are shown in Figure 7. The training A_z values of the LDA classifier on the two training subsets were 0.87 ± 0.02 and 0.88 ± 0.01 , respectively. The classifiers achieved A_z values of 0.89 ± 0.02 and 0.85 ± 0.02 on the independent test subsets, respectively. Figure 8 shows the FROC curves for the two test subsets after FP reduction with the corresponding trained LDA classifiers. An average FROC curve was derived from these two FROC curves by averaging the FP/images at the corresponding sensitivities. This average test FROC curve is plotted in Figure 9 for comparison with the other FROC curves, described next.

In addition to using the mass data set containing 110 cases for the cross validation training and testing, we used a no-mass data set containing 90 cases with 180 images to evaluate the FP detection rate in normal cases. Since two sets of trained parameters were acquired as a result of the cross validation training, we applied the two trained CAD systems separately to the no-mass data set for FP detection. The number of FP marks produced by the algorithm was determined by counting the detected objects on these normal cases only. The mass detection sensitivity was determined by counting only the abnormal objects on each of the test mass subsets. The combination of the sensitivity from each of the test mass subsets and the FP rate

from the normal data set at the corresponding detection thresholds resulted in a test FROC curve.

The two test FROC curves were then averaged, as described above, to obtain an overall FROC curve quantifying the test performance of the CAD system. Figures 9(a) and 9(b) show the comparison of the average FROC curves with the FP rates estimated from the two data sets. The test FROC curve with the FP rate estimated from the no-mass data set showed a case-based detection sensitivity of 70%, 80%, and 90% at 0.85, 1.31, and 2.14 FP marks/image, which are slightly higher than the FP rates of 0.7, 1.1, and 1.8 marks/image, respectively, estimated from the mass data set. Since our mass detection algorithm limits the maximum number of output marks to be 3 at the final stage, the FP marker rates will be slightly higher if the detection is performed in no-mass images. However, many images do not reach the maximum of 3 marks so that the difference in the FP marker rate between the mass and no-mass set is less than one. We also analyzed the detection accuracy of the system for malignant and benign masses separately. Figures 10(a) and 10(b) show the average FROC curves for detection of malignant and benign masses.

The average test FROC curves of the CAD system using the GE processed images as input were compared to those of the CAD system using raw images as input and Laplacian pyramid multiscale preprocessing as shown in Figure 9. The FROC curves were plotted as the detection sensitivity as a function of the number of FP marks per image on the mass data set.

The CAD system using the GE processed images as input achieved a case-based sensitivity of 70%, 80%, and 90% at 0.9, 1.6, and 3.1 FP marks/image, respectively, compared with 0.7, 1.1, and 1.8 FP marks/image on the CAD system using raw images as input.

IV. DISCUSSION

Several FFDM systems have been approved for clinical applications. It is important to develop a CAD system that can easily be adapted to images acquired by FFDM systems from different manufacturers. In this study, we are developing a CAD system that uses the raw FFDMs as the input. Since digital detectors generally have a linear response to x-ray exposure, the raw pixel values are a linear function of the absorbed x-ray energy in the detector. The signal range between different digital detectors can therefore be normalized linearly with respect to each other. Although the spatial resolution and noise properties of the images from different detectors are still different, the use of raw images already reduces one of the major differences between mammograms from different FFDM systems. For preprocessing of the raw images, we developed a multi-resolution enhancement method. An example of a typical mammogram processed by the GE method and our method is compared in Figure 5. As seen from this example, the enhancement of mammographic structures was stronger for our processed image than for the GE processed image. From a comparison of their histograms, it was found that the two histograms are very similar except for the average gray level.

For the evaluation of the effect of the preprocessing methods on computerized mass detection, we observed that our Laplacian pyramid preprocessing method provided higher detection accuracy than the GE processing method. As shown in Fig. 5, the Laplacian pyramid preprocessing method applies a stronger edge enhancement to the image than the GE method. Our preprocessing method aims at enhancing the image structures for computer vision whereas the GE processing method was designed to enhance the image for human visual interpretation. The stronger enhancement used for preprocessing the raw images appeared to improve the accuracy of the computer in detecting the masses.

Currently, there is no established statistical analysis method for testing the significance of the difference between two FROC curves generated by a CAD system. Chakraborty et al. proposed to use an alternative free-response ROC (AFROC) method⁴⁹ to transform the FROC data to AFROC data, to which the curve fitting software and statistical significance tests for ROC analysis can then be applied and demonstrated its application to human observer performance rating data. In the AFROC method, false-positive images (FPI) instead of FPs per image are counted. The confidence rating of an FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions are made on the same image. We applied the AFROC method to evaluate the differences in pairs of our FROC curves that used the no-mass set for estimation of the FP rates. The ROCKIT software developed by

Metz et al.⁴⁷ was used to analyze the AFROC data. The comparison of A_I and p values is summarized in Table 1. The area under the fitted AFROC curve (A_I) was 0.44 and 0.39, respectively, on mass test subset 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.37 and 0.31, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the fitted AFROC curve for our processed images and that for the GE processed images was statistically significant ($p < 0.05$) for both test subsets. However, all four fitted AFROC curves deviated systematically from the AFROC data (see two examples plotted in Figure 11 for the test subset 1). It is uncertain whether the AFROC method is applicable to our FROC data and thus whether the statistical significance testing is valid.

More recently, Chakraborty et al.⁵⁰ proposed a JAFROC method and provided software to estimate the statistical significance of the difference between two FROC curves. We also applied the JAFROC analysis to the two pairs of FROC curves. The figure-of-merit (FOM) from the output of the JAFROC software was 0.46 and 0.41, respectively, on mass test subset 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.39 and 0.34, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the FOM for our processed images and that for the GE processed images was again statistically significant ($p < 0.05$). The FOM values

were about 0.02 higher than the corresponding A_1 values. The JAFROC software did not provide a fitted curve or a goodness-of-fit indicator in the output so that it is not known whether this model fits our FROC data better than the AFRPC method. Although both methods indicate that the improvement in the FROC performance using our Laplacian pyramid processed images is statistically significant, further investigations are needed to study whether these models are valid for analyzing the FROC performance of CAD systems.

The prescreening technique is an important task in a CAD system. A number of researchers have developed methods for detection of suspicious masses on SFMs and CRs. The previous methods produced between 10 to 30 FPs/image for a mass detection sensitivity of approximately 90%. However, it is difficult to compare the effectiveness of the different methods because of the differences in the image recording systems and in the data sets. In this study, we developed a new method that combines gradient field information, which was originally developed for the detection of lung nodules on chest x-ray images⁴³, and gray level information⁴⁴ for prescreening mass candidates on the FFDMs. The new method produced 18.9 objects/image at 93% sensitivity in the prescreening step, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone.

The texture features in this study were extracted by using the SGLD matrix. A total of 572 features were included in our initial feature pool. These features were also used by our

CAD system previously developed for SFMs. An average number of 19.5 features were selected by using a stepwise feature selection method. The A_z values for the LDA classifiers were 0.87 ± 0.02 and 0.88 ± 0.01 on the two training subsets, and 0.89 ± 0.02 and 0.85 ± 0.02 on the test subsets, respectively. The slightly higher test A_z from the first test subset than the A_z from its training subset may indicate that some relatively easy cases were assigned, by chance, to that test set during random partitioning. We also investigated if other features could improve the performance of our CAD system. The different feature spaces that we examined included features extracted from principal component analysis applied to the ROI image, run length statistics texture features extracted from the ROI images, and combination of one or both of these feature spaces with the SGLD feature space. However, the test results showed that an LDA classifier designed in the SGLD feature space alone provided the best performance. Although this was found to be true for both our CAD mass detection system for SFMs developed previously and the current system for FFDMs, it is still difficult to conclude that the SGLD features are the best feature set for classification between breast masses and normal tissues. One major concern of the SGLD feature space is that the dependence of the feature values on the pixel pair distance and angular direction leads to a feature pool with a large number of features. Some features in such a large feature space may provide good performance in classification of masses and normal structures by chance. We attempted to alleviate this problem by using an

independent test set to evaluate the classifier performance. However, since we chose the overall system parameters with the knowledge of the performance for the test sets, the evaluation would still amount to validation rather than true testing. We have verified that our CAD system for SFMs can achieve reasonable performance in a true independent data set³⁶ and a prospective pilot clinical trial¹⁶. The performance of the current CAD system for FFDMs will have to be evaluated similarly when independent data sets become available.

The detection performance of a CAD system for malignant masses is more important than its performance for all masses. Figures 10(a) and 10(b) indicate that the sensitivity of the system is higher for malignant masses than for benign masses. This is consistent with our observation in previous studies of our CAD system for digitized SFMs³⁶. However, since our current data set contained only 23 malignant cases, there will be large statistical uncertainty in the evaluation of sensitivity in this subset. A larger data set is being collected for comparing the detection performances of the CAD system between malignant and benign masses and also for the purpose of classifying malignant and benign masses. Furthermore, CAD algorithms developed for SFMs have been proven to be useful as a second opinion to assist radiologists in mammographic interpretation. Because of the higher SNR and linear response of digital detectors, there is also a potential that FFDMs can improve the sensitivity of breast cancer detection, especially in dense breasts. Several studies have been or are being conducted to

compare FFDM with SFM in screening cohorts. It is also important to compare the performance of CAD systems between FFDMs and SFMs. A study is underway to compare the performance of the two systems on pairs of FFDM and SFM obtained from the same patients⁵¹.

V. CONCLUSION

Several FFDM systems have been approved for clinical applications. It is important to develop CAD systems for breast cancer detection in FFDM. In this work, we developed a CAD system that uses the raw FFDMs as the input. A multiresolution Laplacian pyramid enhancement method was devised to preprocess the raw FFDMs. A new prescreening method that combined gradient field analysis with gray level information was developed to identify mass candidates. Rule-based and LDA classifiers in a feature space which consisted of morphological features and SGLD texture features were designed to differentiate masses from normal tissues. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% with an estimate of 0.85, 1.31, and 2.14 FP marks/image, respectively, on normal cases. The results indicate that our mass detection CAD scheme can be useful for detecting masses on FFDMs. Studies are underway to further optimize the processing parameters, the feature extraction, and the classifiers for FP reduction. Comparison of mass detection performance of our CAD system for FFDMs and that for SFMs is also in progress.

Acknowledgments

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD 17-02-1-0214 and W81XWH-04-1-0475. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC and ROCKIT programs.

Table 1. Estimation of the statistical significance in the difference between the FROC performance of the CAD system using the FFDM raw images as input and processed with our Laplacian pyramid method and that of the CAD system using GE processed images as input. The FROC curves with the FP rates obtained from the no-mass data set (Figure 9) were compared.

| | A_I (AFROC) | | | FOM (JAFROC) | | |
|--------------------|---------------|---------------|----------|---------------|---------------|----------|
| | Test subset 1 | Test subset 2 | p values | Test subset 1 | Test subset 2 | p values |
| Raw + LP processed | 0.44 | 0.39 | 0.012 | 0.46 | 0.41 | 0.006 |
| GE processed | 0.37 | 0.31 | 0.0009 | 0.39 | 0.34 | 0.012 |

References

¹"American Cancer Society, www.cancer.org 2004, "Statistics for 2004","

²C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," *Cancer* 75, 1619-1626 (1995).

³S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," *Am J Roentgenol* 171, 29-33 (1998).

⁴B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," *Cancer* 91, 1699-1703 (2001).

⁵B. J. Hillman, L. L. Fajardo, T. B. Hunter, B. Mockbee, C. E. Cook, R. M. Hagaman, J. C. Bjelland, C. S. Frey, and C. J. Harris, "Mammogram interpretation by physician assistants," *Am J Roentgenology* 149, 907-911 (1987).

⁶C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists - Findings from a national sample," *Archives of Internal Medicine* 156, 209-213 (1996).

⁷R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology* 219, 192-202 (2001).

⁸J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," *J. National Cancer Institute* 95, 1384-1393 (2003).

⁹F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National Cancer Institute perspective," Orlando, Florida,

¹⁰C. J. Vyborny, "Can computers help radiologists read mammograms?," *Radiology* 191, 315-317 (1994).

¹¹H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," *Invest. Radiol.* 25, 1102-1110 (1990).

¹²L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* 215, 554-562 (2000).

¹³T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* 220, 781-786 (2001).

¹⁴R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," *Am J Roentgenology* 181, 687-693 (2003).

¹⁵S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can Computer-aided Detection with Double Reading of Screening Mammograms Help Decrease the False-Negative Rate? Initial Experience," *Radiology* 232, 578-584 (2004).

¹⁶M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, et al., "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - A pilot clinical trial," *Radiology* 231, 208-214 (2004).

¹⁷D. Gur, J. H. Sumkin, and H. E. Rockette, "Response to Re: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," *J National Cancer Institute* 96, 1261 (2004).

¹⁸S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," *J National Cancer Institute* 96, 1260-1261 (2004).

¹⁹F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis.," Radiology 89, 211-115 (1967).

²⁰J. L. Semmlow, A. Shadagopappan, L. V. Ackerman, W. Hand, and F. S. Alcorn, "A fully automated system for screening mammograms," Comput. Biomed. Res. 13, 350-362 (1980).

²¹T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach," Computers and Biomedical Research 24, 273-295 (1991).

²²F. F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," Med. Phys. 18, 955-963 (1991).

²³C. Kimme, B. J. O'Laughlin, and J. Sklansky, *Automatic detection of suspicious abnormalities in breast radiographs*, Data Structures, Computer Graphics and Pattern Recognition (Academic Press, New York, 1975).

²⁴W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," Radiology 191, 331-337 (1994).

²⁵S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," Comput. Biomed. Res. 25, 218-237 (1992).

²⁶N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," IEEE Trans. Med. Img. 15, 611-619 (1996).

²⁷S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," IEEE Trans. Med. Img. 8, 377-386 (1989).

²⁸D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," IEEE Trans. Med. Img. 9, 233-241 (1990).

²⁹A. F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," IEEE Trans. Med. Img. 13, 725-740 (1994).

³⁰N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," *Med. Phys.* 23, 1685-1696 (1996).

³¹B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," *Acad. Radiol.* 2, 959-966 (1995).

³²H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computer detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Img.* 18, 369-378 (1999).

³³L. Li, R. A. Clark, and J. A. Thomas, "Computer-aided diagnosis of masses with full-field digital mammography," *Acad. Radiol.* 9, 4-12 (2002).

³⁴D. Gur, J. S. Stalder, L. A. Hardesty, B. Zheng, J. H. Sumkin, D. M. Chough, B. E. Shindel, and H. E. Rockette, "Computer-aided Detection Performance in Mammographic Examination of Masses: Assessment," *Radiology* 233, 418-423 (2004).

³⁵N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Trans. Med. Img.* 15, 59-67 (1996).

³⁶N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquereault, and L. M. Hadjiiski, "Breast cancer detection: Evaluation of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," *Radiology* 224, 217-224 (2002).

³⁷P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications* COM-31, 337-345 (1983).

³⁸S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale contrast enhancement for radiographies: laplacian pyramid versus fast wavelet transform," *IEEE Trans. Med. Img.* 21, 343-353 (2002).

³⁹N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. System, Man, Cybernetics* 9, 62-66 (1979).

⁴⁰A. Burgess, "On the noise variance of a digital mammography system," *Med.Phys* 31, 1987-1995 (2004).

⁴¹H. P. Chan, S. C. B. Lo, L. T. Niklason, D. M. Ikeda, and K. L. Lam, "Image compression in digital mammography: Effects on computerized detection of subtle microcalcifications," *Med. Phys.* 23, 1325-1336 (1996).

⁴²H. Kobatake and S. Hashimoto, "Convergence Index Filter for Vector Fields," *IEEE Trans. Img. Proc.* 8, 1029-1038 (1999).

⁴³J. Wei, Y. Hagihara, and H. Kobatake, "Detection of rounded opacities on chest radiographs using convergence index filter," *ICIAP 99*, Venice, September 27-29, 757 - 761 (1999).

⁴⁴N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," *Med. Phys.* 26, 1642-1654 (1999).

⁴⁵D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis," *Med. Phys.* 24, 903-914 (1997).

⁴⁶R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. Sys. Man. and Cybern. SMC-3*, 610-621 (1973).

⁴⁷C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Stat. Med.* 17, 1033-1053 (1998).

⁴⁸B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size.," *Med. Phys.* 27, 1509-1522 (2000).

⁴⁹D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Med. Phys.* 16, 561-568 (1989).

⁵⁰D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: modeling, analysis, and validation," *Medical Physics* 31, 2313-2330 (2004).

⁵¹J. Wei, B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, and M. A. Helvie, "Computer aided diagnosis system for mass detection: comparison of performance on full-field digital mammograms and digitized film mammograms," *RSNA 2003*, Chicago, November 30-December 5 387 (2003).

Figure Captions

Figure 1. The information of our mass data set: (a) distribution of mass sizes, (b) distribution of mass shapes, (c) distribution of mass margins, C: circumscribed, Ind: indistinct, M: microlobulated, Ob: obscured, Sp: spiculated, (d) distribution of the breast density in terms of BI-RADS category estimated by an MQSA radiologist.

Figure 2. Schematic diagram of our CAD system for mass detection on FFDM. The system is developed for screening mammography so that all masses, regardless of malignant or benign, are considered positive. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

Figure 3. Schematic diagram for the image preprocessing stage of our mass detection system, which includes breast boundary segmentation, logarithmic image transformation, and Laplacian pyramid multiscale enhancement.

Figure 4. Multiscale enhancement using the Laplacian pyramid decomposition method: Laplacian decomposition tree on the left side and the Gaussian reconstruction tree on the right side. The different levels of the Gaussian pyramid images are denoted by G_b ($i=0, \dots, n$). The error images at different levels of the Laplacian pyramid are denoted by L_b ($i=0, \dots, n$). The primed quantities G_i' and L_i' denoted the images at different levels after enhancement. Σ denotes the summation operation. The image

is downsampled by a factor of 2 when it goes down every level of the decomposition tree, and upsampled by a factor 2 when it moves up every level of the reconstruction tree.

Figure 5. An example of (a) GE raw image, (b) GE processed image, and (c) our processed image by using the Laplacian pyramid multiscale method. The gray level histogram of each image is also shown. The GE raw image has 14-bit gray levels but the histogram only plotted the lower 12 bits because very few pixels had gray levels higher than 4095.

Figure 6. An example demonstrating the processing steps with our CAD system: (a) object locations identified in prescreening, (b) identified suspicious objects, (c) detected objects after FP reduction, and (d) image superimposed with ROIs identifying the detected objects. The true mass is indicated by an arrow.

Figure 7. The test ROC curves from the two independent mass subsets. The LDA classifiers using text features achieved an A_z value of 0.89 ± 0.02 for test subset 1 and 0.85 ± 0.02 for test subset 2 in the classification of mass and normal breast tissues.

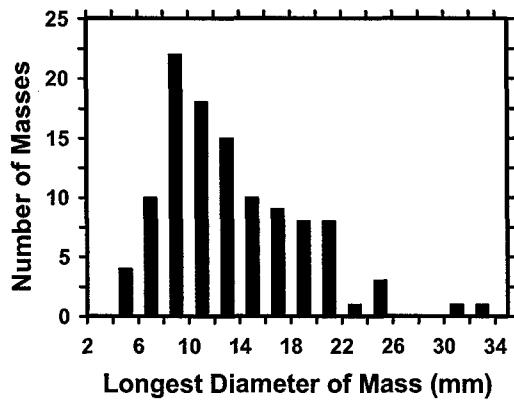
Figure 8. The test FROC curves from two independent mass data sets for the CAD system using the raw images as input and processed with the Laplacian pyramid method. The FP rate was estimated from the mammograms with masses. (a) Image-based

FROC curves, (b) Case-based FROC curves.

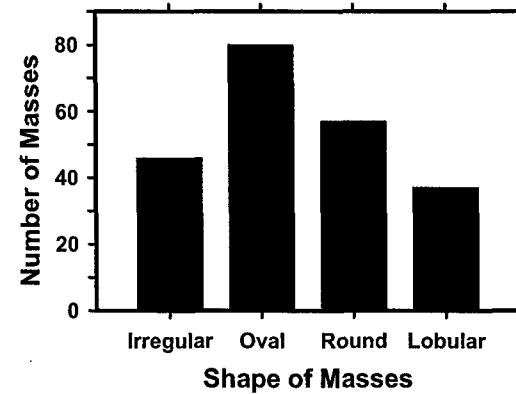
Figure 9. Comparison of the average test FROC curves obtained from: (1) the CAD system using raw images as input, with the FP rate estimated from the mammograms with masses, (2) the CAD system using raw images as input, with the FP rate estimated from the normal mammograms without masses, and (3) the CAD system using GE processed images as input, with the FP rate estimated from the GE processed mammograms with masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

Figure 10. Comparison of the average test FROC curves for the malignant and benign mass sets. The CAD system using raw images as input was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

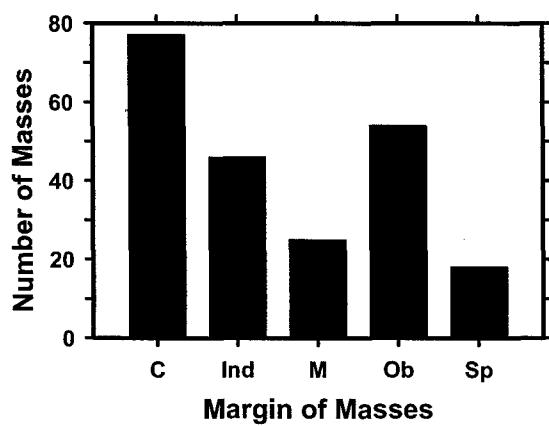
Figure 11. Comparison of alternative free-response receiver operating characteristic (AFROC) curves. The raw curves were transformed from the FROC curves of mass detection on test subset one using either the raw images as input and processed with the Laplacian pyramid method (LP) or the GE processed images as input. The FP rate was estimated from the mammograms without masses. The fitted AFROC curves were obtained by applying the ROCKIT program to the transformed AFROC data.



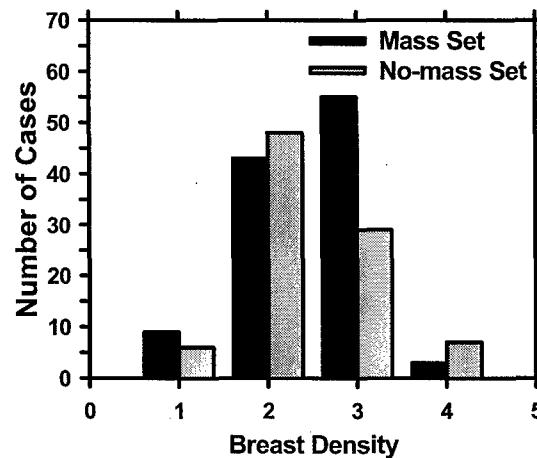
(a)



(b)



(c)



(d)

Figure 1. The information of our mass data set: (a) distribution of mass sizes, (b) distribution of mass shapes, (c) distribution of mass margins, C: circumscribed, Ind: indistinct, M: microlobulated, Ob: obscured, Sp: spiculated, (d) distribution of the breast density in terms of BI-RADS category estimated by an MQSA radiologist.

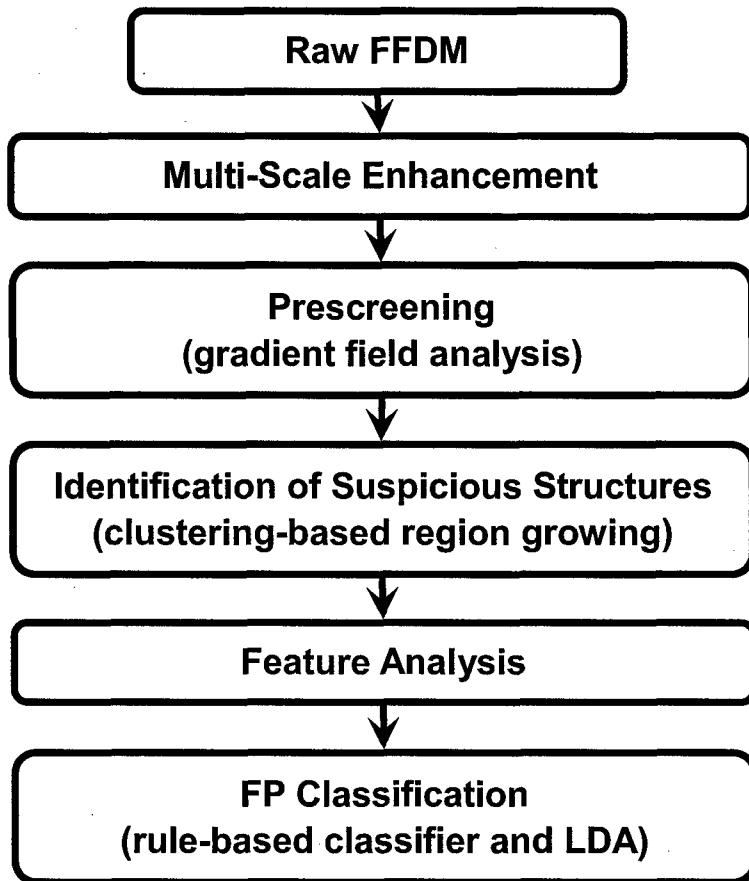


Figure 2. Schematic diagram of our CAD system for mass detection on FFDM. The system is developed for screening mammography so that all masses, regardless of malignant or benign, are considered positive. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

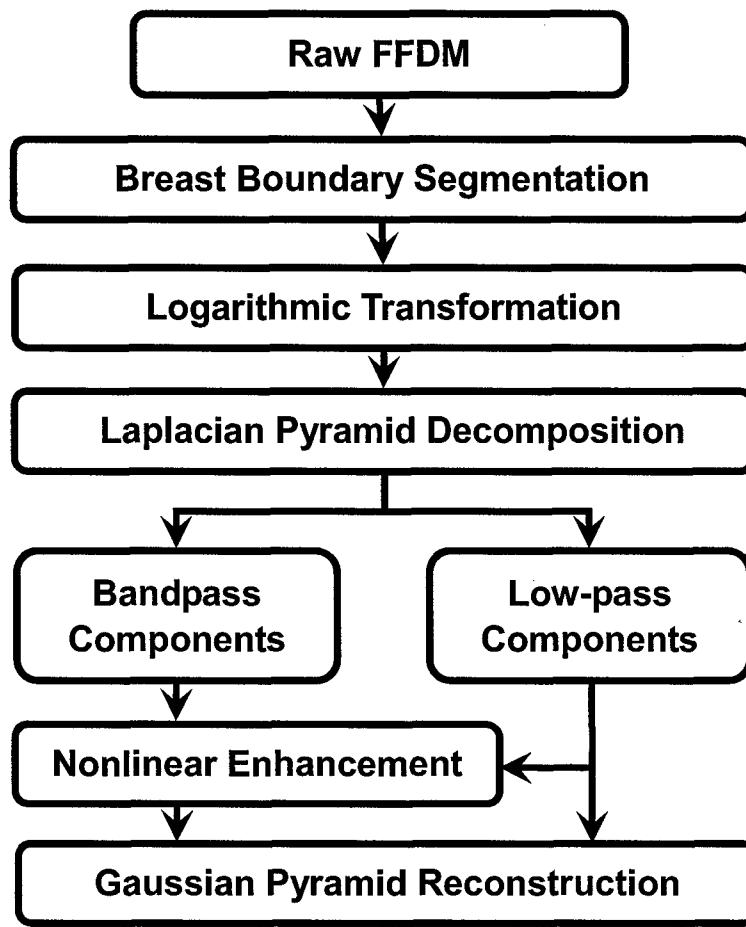


Figure 3. Schematic diagram for the image preprocessing stage of our mass detection system, which includes breast boundary segmentation, logarithmic image transformation, and Laplacian pyramid multiscale enhancement.

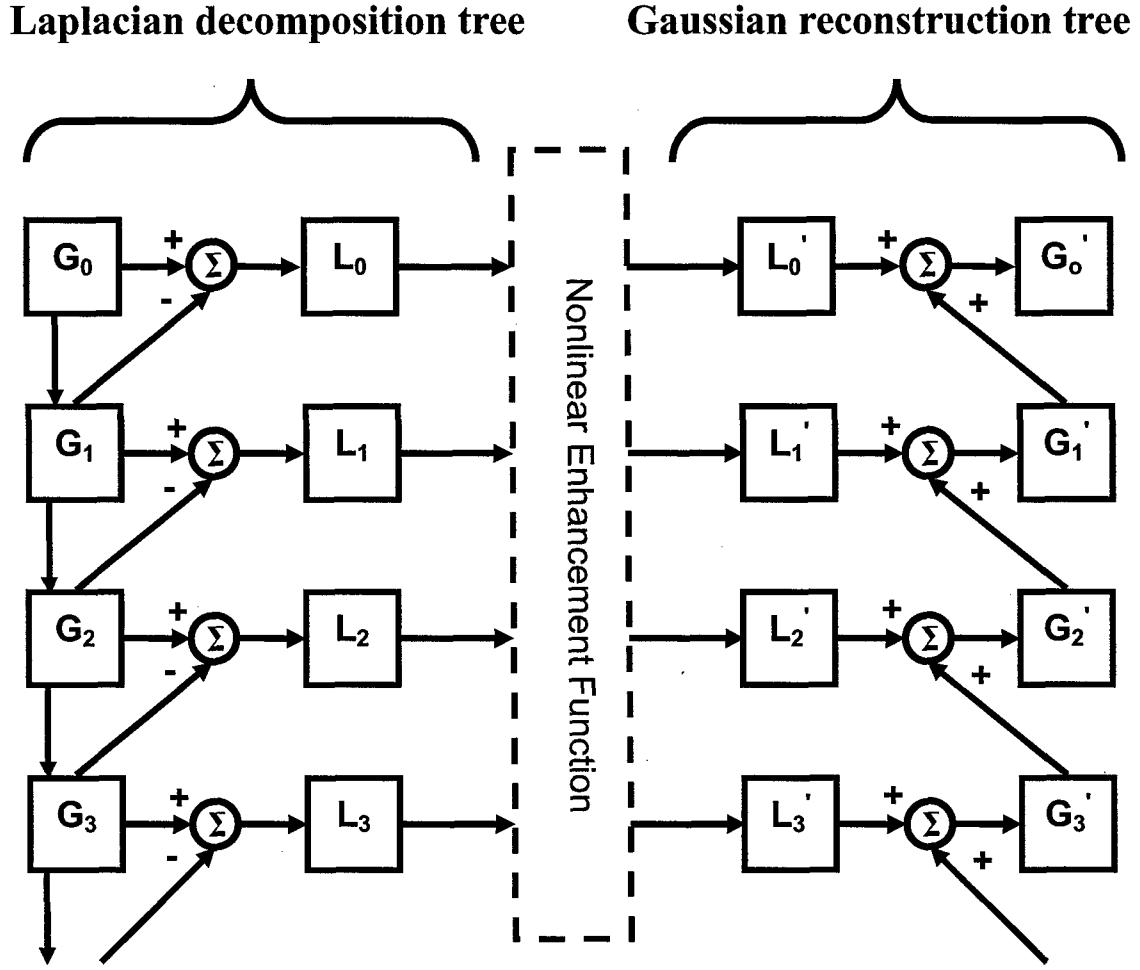
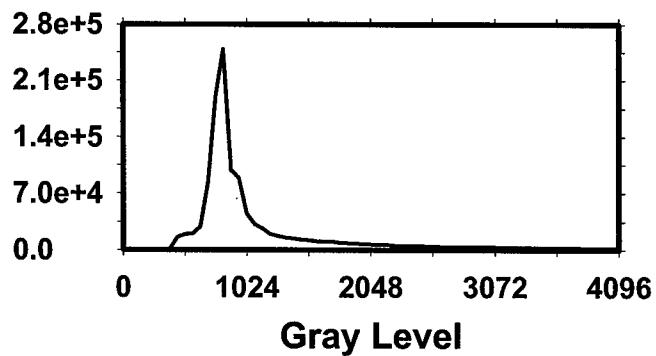
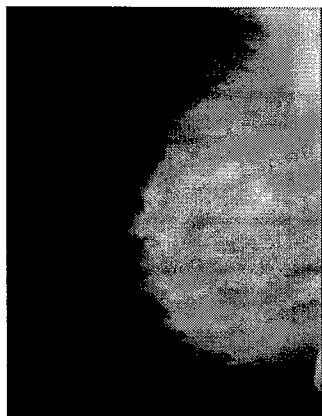
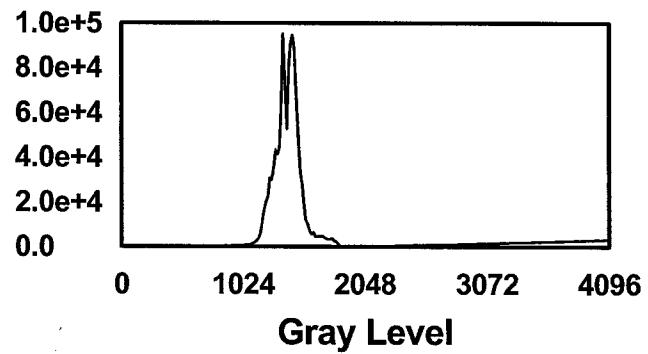
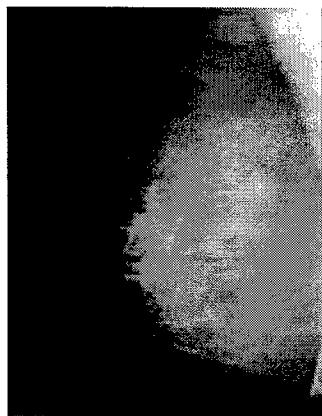


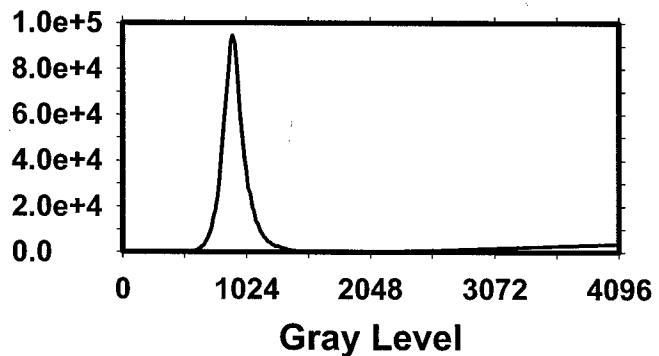
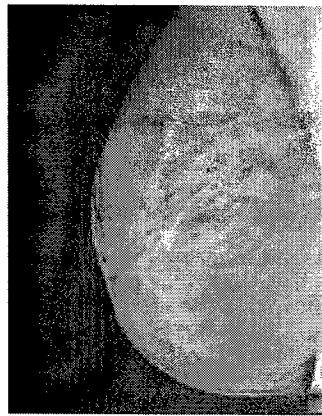
Figure 4. Multiscale enhancement using the Laplacian pyramid decomposition method: Laplacian decomposition tree on the left side and the Gaussian reconstruction tree on the right side. The different levels of the Gaussian pyramid images are denoted by G_i ($i=0, \dots, n$). The error images at different levels of the Laplacian pyramid are denoted by L_i ($i=0, \dots, n$). The primed quantities G'_i and L'_i denote the images at different levels after enhancement. Σ denotes the summation operation. The image is downsampled by a factor of 2 when it goes down every level of the decomposition tree, and upsampled by a factor 2 when it moves up every level of the reconstruction tree.



(a)

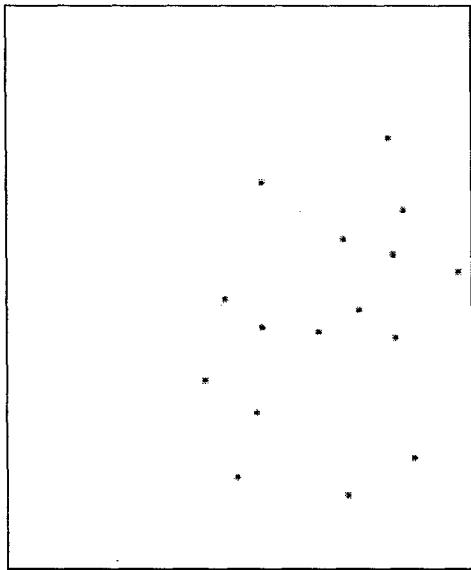


(b)

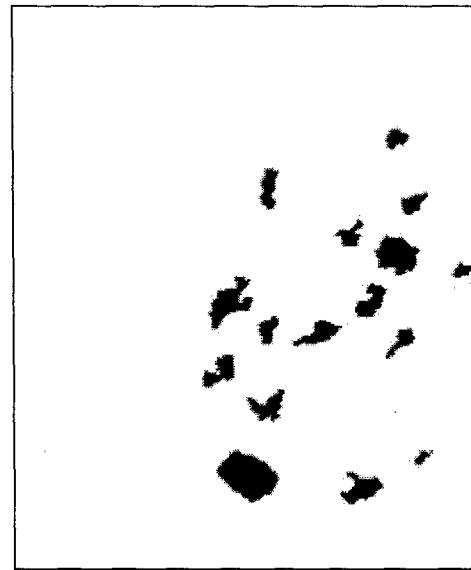


(c)

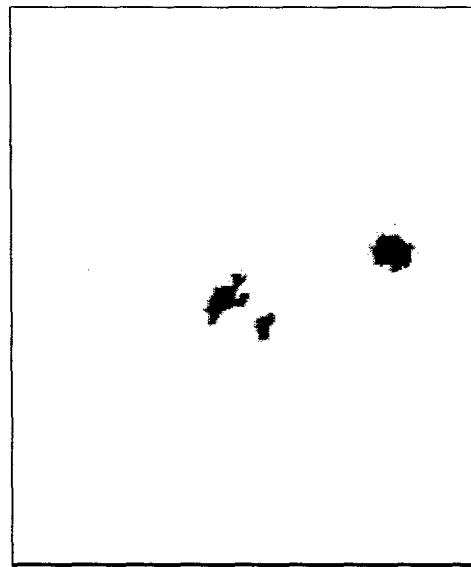
Figure 5. An example of (a) GE raw image, (b) GE processed image, and (c) our processed image by using the Laplacian pyramid multiscale method. The gray level histogram of each image is also shown. The GE raw image has 14-bit gray levels but the histogram only plotted the lower 12 bits because very few pixels had gray levels higher than 4095.



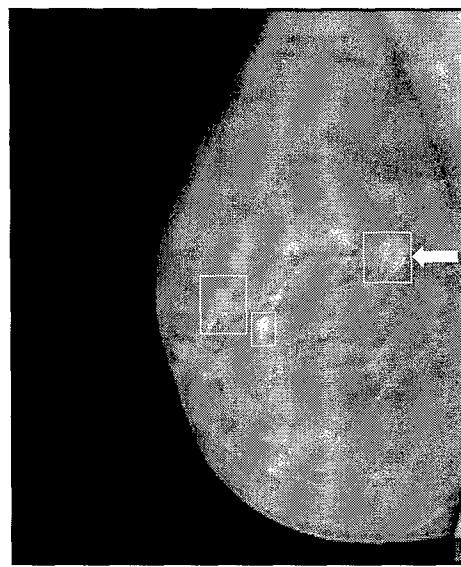
(a)



(b)



(c)



(d)

Figure 6. An example demonstrating the processing steps with our CAD system: (a) object locations identified in prescreening, (b) identified suspicious objects, (c) detected objects after FP reduction, and (d) image superimposed with ROIs identifying the detected objects. The true mass is indicated by an arrow.

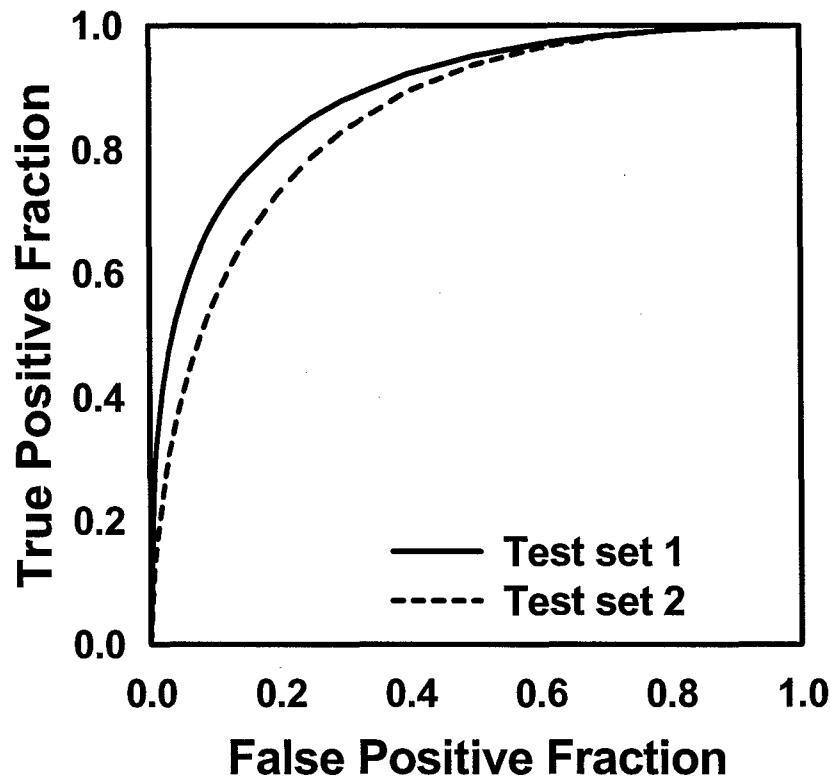
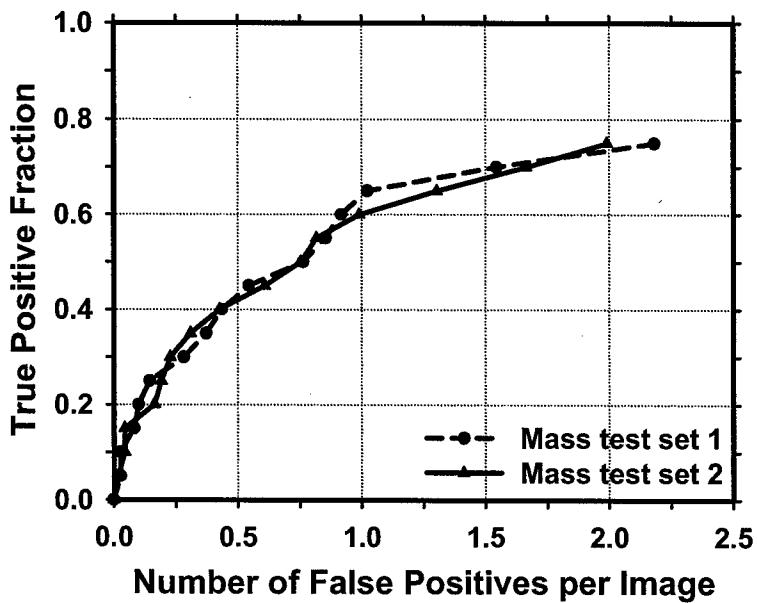
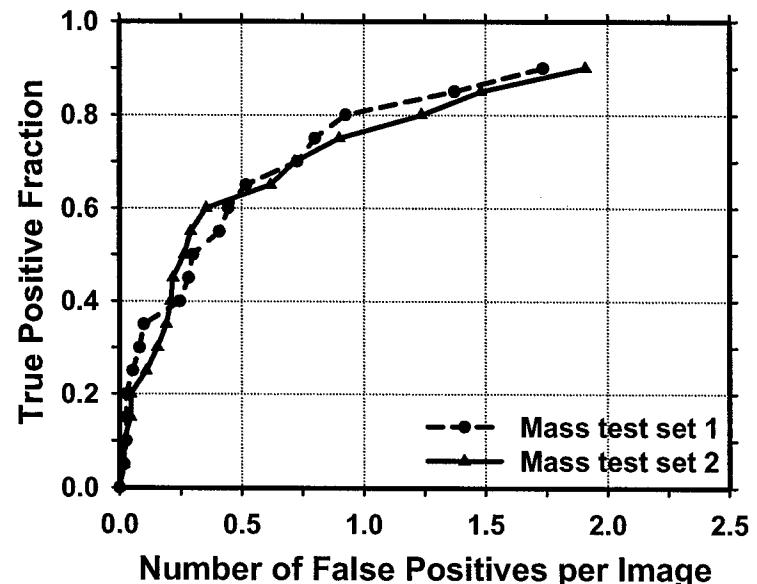


Figure 7. The test ROC curves from the two independent mass subsets. The LDA classifiers using text features achieved an A_z value of 0.89 ± 0.02 for test subset 1 and 0.85 ± 0.02 for test subset 2 in the classification of mass and normal breast tissues.

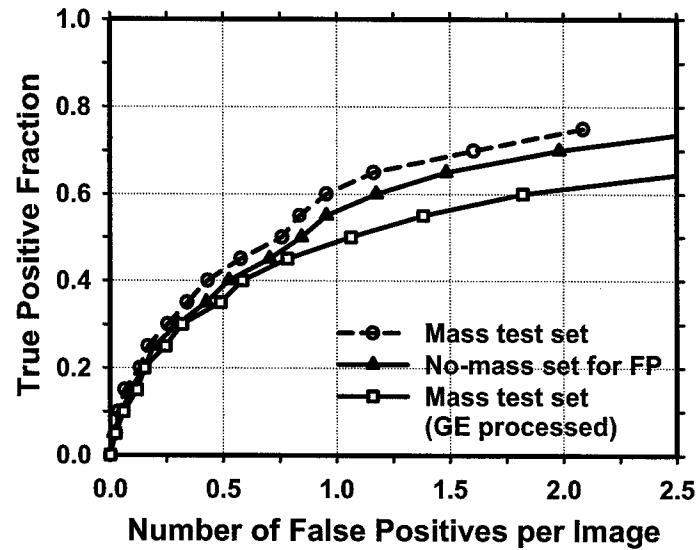


(a)

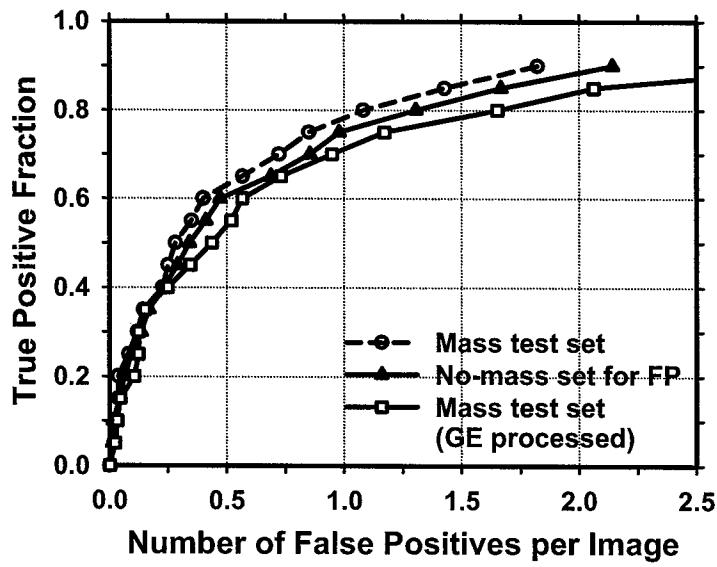


(b)

Figure 8. The test FROC curves from the two independent mass subsets for the CAD system using the raw images as input and processed with the Laplacian pyramid method. The FP rate was estimated from the mammograms with masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

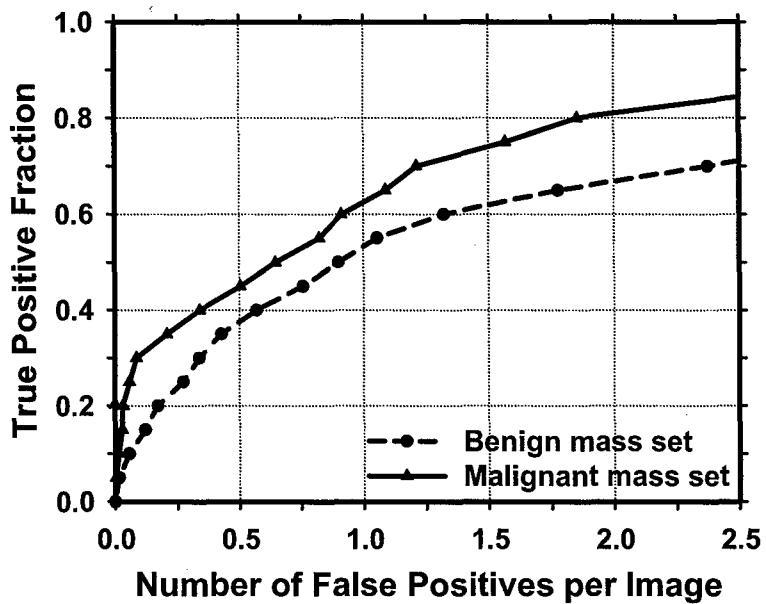


(a)

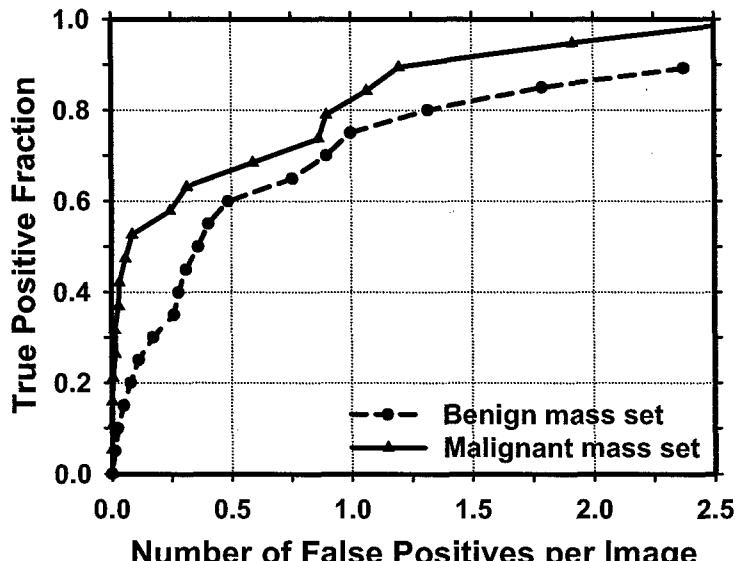


(b)

Figure 9. Comparison of the average test FROC curves obtained from: (1) the CAD system using raw images as input, with the FP rate estimated from the mammograms with masses, (2) the CAD system using raw images as input, with the FP rate estimated from the normal mammograms without masses, and (3) the CAD system using GE processed images as input, with the FP rate estimated from the GE processed mammograms with masses. (a) Image-based FROC curves, (b) Case-based FROC curves.



(a)



(b)

Figure 10. Comparison of the average test FROC curves for the malignant and benign mass sets. The CAD system using raw images as input was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

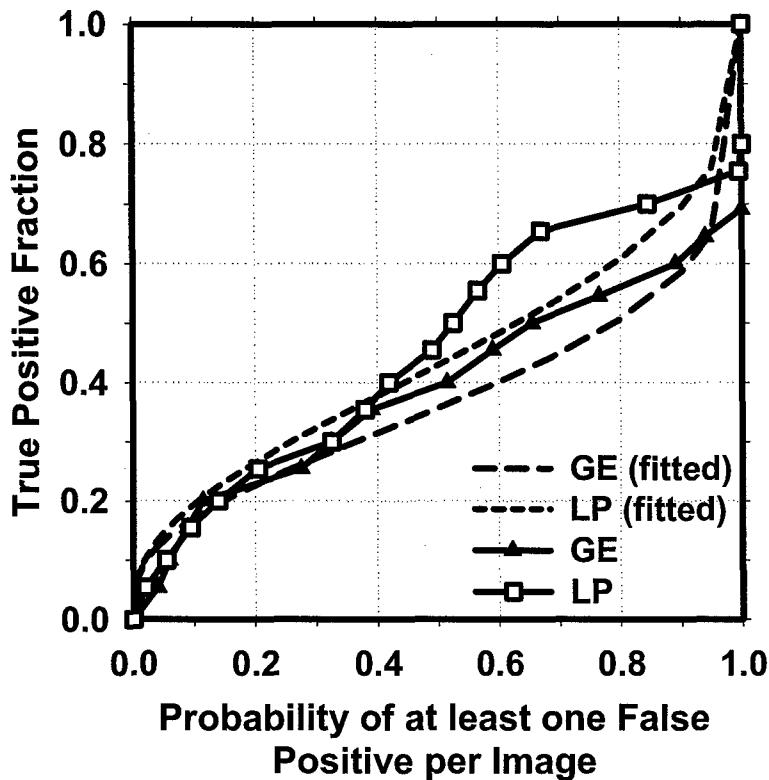


Figure 11. Comparison of alternative free-response receiver operating characteristic (AFROC) curves. The raw curves were transformed from the FROC curves of mass detection on test subset 1 using either the raw images as input and processed with the Laplacian pyramid method (LP) or the GE processed images as input. The FP rate was estimated from the mammograms without masses. The fitted AFROC curves were obtained by applying the ROCKIT program to the transformed AFROC data.

Computer-aided detection of breast masses on mammograms: performance improvement using a dual system

Jun Wei^{*a}, Berkman Sahiner^a, Lubomir M. Hadjiiski^a, Heang-Ping Chan^a, Mark A. Helvie^a, Marilyn A. Roubidoux^a, Nicholas Petrick^b, Jun Ge^a, Chuan Zhou^a

^aDepartment of Radiology, University of Michigan, Ann Arbor

^bCenter of Devices and Radiological Health, U. S. Food and Drug Administration, Rockville, MD

ABSTRACT

We have developed a computer-aided detection (CAD) system for breast masses on mammograms. In this study, our purpose was to improve the performance of our mass detection system by using a new dual system approach which combines a CAD system optimized with "average" masses with another CAD system optimized with subtle masses. The latter system is trained to provide high sensitivity in detecting subtle masses. For an unknown mammogram, the two systems are used in parallel to detect suspicious objects. A feed-forward backpropagation neural network trained to merge the scores of the two linear discriminant analysis (LDA) classifiers from the two systems makes the final decision in differentiation of true masses from normal tissue. A data set of 86 patients containing 172 mammograms with biopsy-proven masses was partitioned into a training set and an independent test set. This data set is referred to as the average data set. A second data set of 214 prior mammograms was used for training the second CAD system for detection of subtle masses. When the single CAD system trained on the average data set was applied to the test set, the A_z for false positive (FP) classification was 0.81 and the FP rates were 2.1, 1.5 and 1.3 FPs/image at the case-based sensitivities of 95%, 90% and 85%, respectively. With the dual CAD system, the A_z was 0.85 and the FP rates were improved to 1.7, 1.2 and 0.8 FPs/image at the same case-based sensitivities. Our results indicate that the dual CAD system can improve the performance of mass detection on mammograms.

Keywords: computer-aided detection (CAD), mass detection, dual CAD system

1. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 to 55 years of age¹. It has been reported that early diagnosis and treatment can improve significantly the chance of survival for patients with breast cancer²⁻⁴. Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially⁵⁻⁷. Computer-aided detection (CAD) is considered to be one of the promising approaches that may improve the sensitivity of detecting early breast cancer in screening mammography. It has been shown that CAD can increase the cancer detection rate by radiologists both in the laboratory and in clinical practice⁸⁻¹³.

We have been developing CAD systems for detection and characterization of mammographic masses and microcalcifications. Detection of masses on mammograms is more challenging than detection of microcalcifications because the normal fibroglandular tissue in the breast causes false positives (FPs) by mimicking masses and causes false negatives due to overlapping with the lesions. Therefore, mass detection systems generally have lower sensitivity and higher FP rate than microcalcification detection systems. In this study, we are investigating the effectiveness of a dual system approach for improving the performance of mass detection on mammograms.

* jvwei@umich.edu, phone: 734-647-8553, CGC B2103, 1500 E. Medical Center Dr., Ann Arbor, MI 48109-0904

2. MATERIALS AND METHODS

2.1 Materials

The data set we used in this study contained 86 cases. Each case included the current mammograms that were obtained before biopsy and the prior mammograms obtained from previous exams. The prior mammograms were used for training the second system because masses on prior mammograms are generally more subtle than those on current mammograms. The subtle mass set does not have to be obtained from the same cases as the average mass set. The current set contained 172 mammograms and the prior set contained 214 mammograms. All data were collected with Institutional Review Board (IRB) approval. The mammograms in this data set were digitized by a Lumiscan laser scanner with a pixel size of $100\mu\text{m} \times 100\mu\text{m}$ and 12 bits per pixel. All of the current cases had two mammographic views: the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral view. There were 86 biopsy-proven masses in this data set. The true locations of the masses were identified by an experienced MQSA radiologist.

2.2 Methods

In order to improve the performance of our CAD system for detection of subtle masses, we developed a new dual system approach which combines a system trained with "average" masses with another system trained with subtle masses. When the trained dual system is applied to an unknown mammogram, the two CAD systems are used in parallel to detect suspicious objects on a single mammogram. No prior mammogram is needed. The additional FPs from the use of two systems are reduced by feature classification in an information fusion stage. Figure 1 shows the block diagram for the dual system.

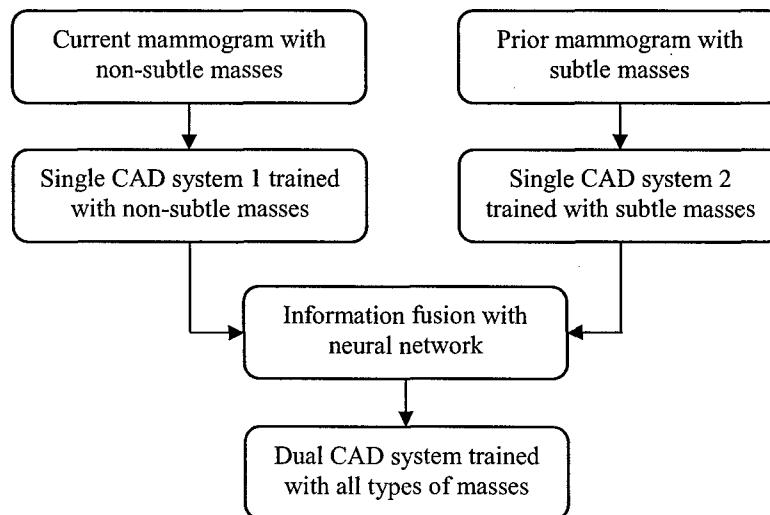


Figure 1. The block diagram of the dual CAD system for mass detection on mammograms.

Our single CAD system consists of five processing steps: 1) digitization, 2) pre-screening of mass candidates, 3) identification of suspicious objects, 4) extraction of feature parameters, and 5) classification between the normal and the abnormal regions by using rule-based and LDA classifiers. The block diagram for the single CAD system is shown in Figure 2. Figure 3 shows an example demonstrating the processing steps with our computer-aided mass

detection system. For the pre-screening stage, we have developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis^{14,15}. The gradient field analysis was used to determine locations of high convergence of radial gradient in the image. A region of interest (ROI) of 256×256 pixels is then identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method¹⁶ in which the location of high gradient convergence is used as the starting point. Figures 3(b) and 3(c) show the initial detection locations and the grown objects, respectively. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. The objects whose new gradient convergence is lower than 80% of the original value are rejected. After prescreening, the suspicious objects are identified by using a clustering-based region growing method. For each suspicious object, eleven morphological features are extracted. Rule-based and LDA classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analysis^{17,18} are performed in each ROI by using the spatial gray level dependence matrices at different pixel spacings and angular directions. In order to obtain the best feature subset and reduce the dimensionality of the feature space to design a robust classifier, feature selection with stepwise linear discriminant analysis was applied. Finally, LDA classification is used to identify potential breast masses. Figure 3(d) shows the final detected objects, and Figure 3(e) shows the locations of these objects superimposed on the mammogram.

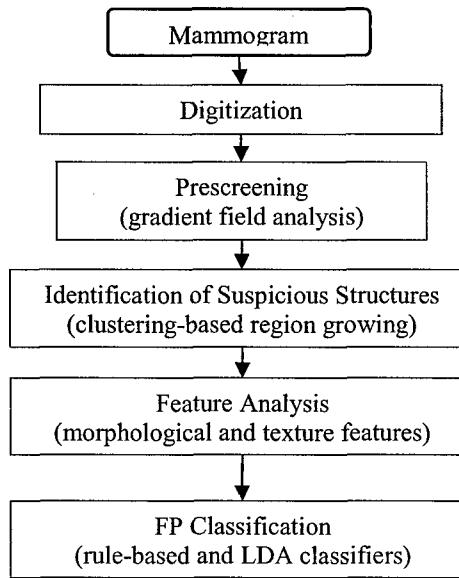


Figure 2. The block diagram of a single CAD system for mass detection on mammograms.

The two single CAD systems were independently trained with the “average” mass set and the subtle mass set, respectively. To merge the information from the two CAD systems, the two LDA discriminant scores from the two CAD systems were used to define a new feature space. A feed-forward backpropagation neural network with 3 hidden nodes was then trained using the LDA feature scores of the training sets as input to differentiate true masses from normal tissue. After the dual CAD system was trained, its performance was evaluated on the independent test set and compared with that of the single CAD system.

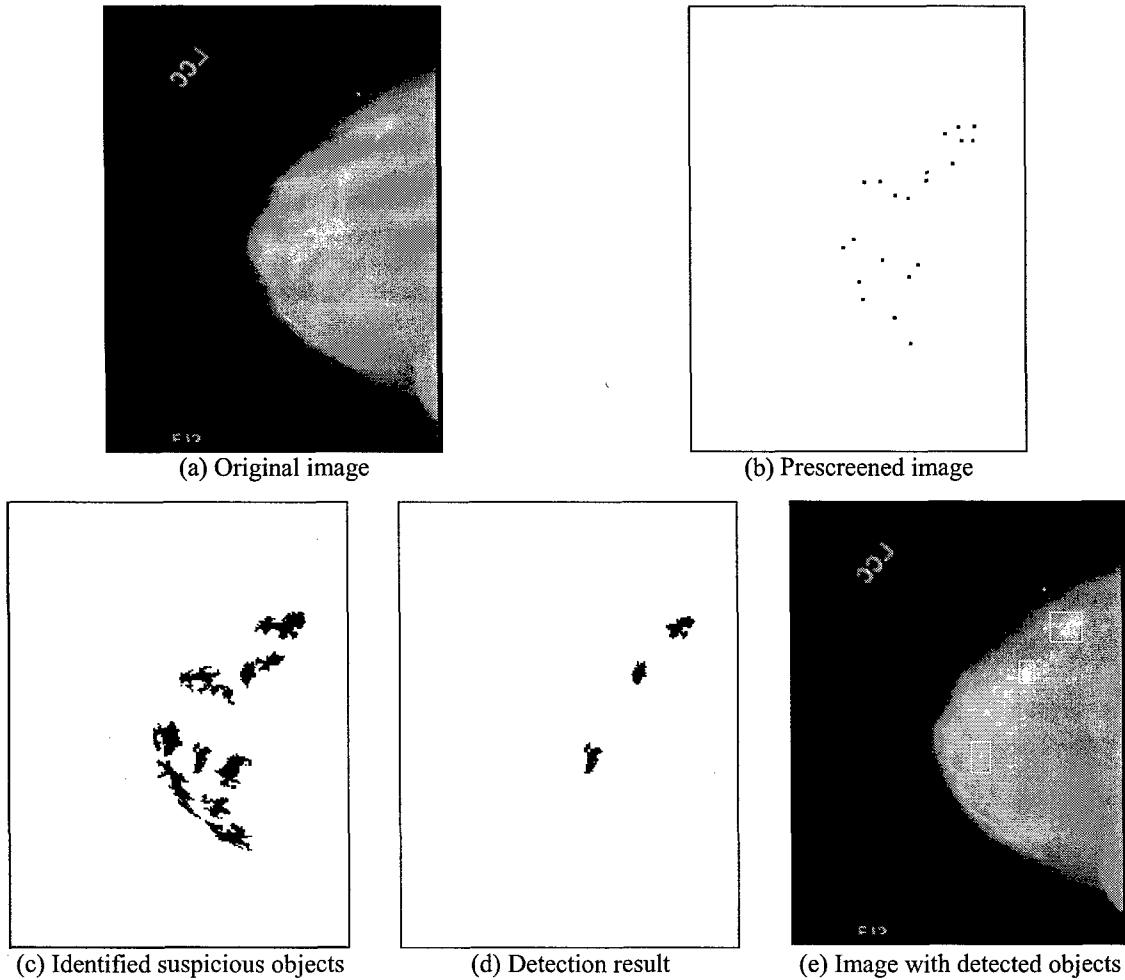


Figure 3. An example demonstrating the processing steps with our single CAD system for mass detection.

3. RESULTS

We randomly separated the cases in our data set into two independent equal sized data sets, each with 43 cases. The training and testing were performed using the cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 86 masses. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and the detection of either or both masses on the two views was considered to be a true-positive (TP); the sensitivity was thus calculated relative to 43 masses. The average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. When the single CAD system trained on the average data set was applied to the test set, the A_z for FP classification was 0.81 and the FPs/image were 2.1, 1.5 and 1.3 at the case-based sensitivities of 95%, 90% and 85%, respectively. With the dual CAD system, the A_z was 0.85 and the FP rates were improved to 1.7, 1.2 and 0.8 FPs/image at the same case-based

sensitivities. Figure 4 and 5 shows the comparison of the test performance of the single and dual CAD systems by using image-based and case-based average FROC curves, respectively.

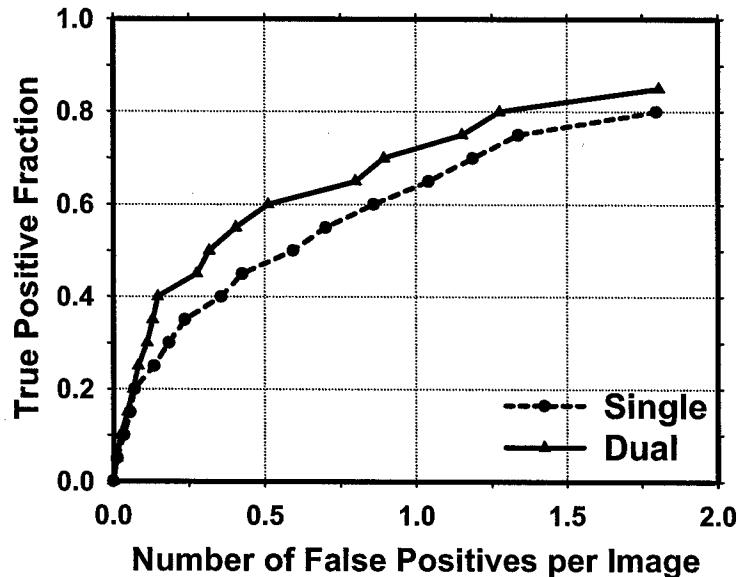


Figure 4. Image-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single: detection by the single CAD system. Dual: detection by the dual CAD system.

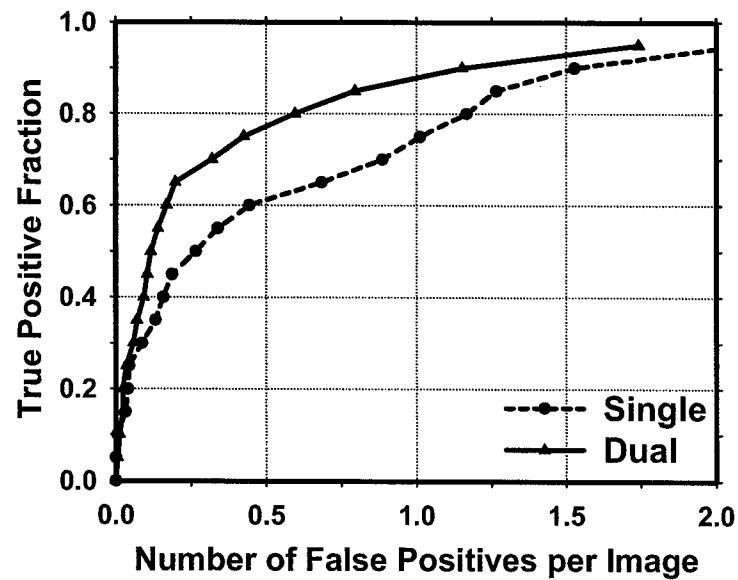


Figure 5. Case-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single: detection by the single CAD system. Dual: detection by the dual CAD system.

4. DISCUSSION AND CONCLUSIONS

We previously developed a CAD system for detection of masses on mammograms. However, we found that it is difficult to train a single system to provide optimal detection for all lesions over the entire spectrum of subtlety. In this study, we developed a dual system which combines a system trained with subtle lesions on prior mammograms and a system trained with masses detected on current mammograms. It was found that the dual CAD system could achieve a higher accuracy than the single CAD system. Further study is underway to optimize the fusion scheme in our dual system.

ACKNOWLEDGMENTS

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD17-03-1-0475 and DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

REFERENCES

1. American cancer society, www.Cancer.Org 2005, "Statistics for 2005"
2. C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
3. S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," *Am J Roentgenol* **171**, 29-33, 1998.
4. B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," *Cancer* **91**, 1699-1703, 2001.
5. C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists - Findings from a national sample," *Archives of Internal Medicine* **156**, 209-213, 1996.
6. R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology* **219**, 192-202, 2001.
7. J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," *J. National Cancer Institute* **95**, 1384-1393, 2003.
8. H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," *Investigative Radiology* **25**, 1102-1110, 1990.
9. L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* **215**, 554-562, 2000.
10. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
11. R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," *Am J Roentgenology* **181**, 687-693, 2003.
12. S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," *Radiology* **232**, 578-584, 2004.

13. M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, et al., "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - A pilot clinical trial," *Radiology* **231**, 208-214, 2004.
14. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou, and Z. Ge, "Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis," *Proc. SPIE Medical Imaging* **5370**, 992-998, 2004.
15. J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical Physics* 2005 (Submitted).
16. N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," *Medical Physics* **26**, 1642-1654, 1999.
17. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis," *Medical Physics* **22**, 1501-1513, 1995.
18. D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis," *Medical Physics* **24**, 903-914, 1997.